

Г.Н. Зайцев

**МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА
в экспериментальной
ботанике**

АКАДЕМИЯ НАУК СССР
ГЛАВНЫЙ БОТАНИЧЕСКИЙ САД

Г.Н. Зайцев

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА в экспериментальной ботанике

Ответственный редактор
член-корреспондент ВАСХНИЛ
В. Н. БЫЛОВ



ИЗДАТЕЛЬСТВО «НАУКА»

МОСКВА 1984

З а й ц е в Г. Н. Математическая статистика в экспериментальной ботанике. М.: Наука, 1984. 424 с.

Книга является одним из самых полных по числу методов справочным биометрическим пособием, доступным для биологов с различной математической подготовкой. Техника расчетов по всем приводимым методам подробно рассматривается на реальных примерах из биологических исследований и практики растениеводства и ботаники. Приведены все справочные таблицы, необходимые для оценки результатов или облегчающие расчеты.

Книга рассчитана на научных работников всех отраслей ботаники и биологии, сотрудников опытных станций, университетов, ботанических садов, заповедников, биологических и медицинских институтов, агрономов, аспирантов и студентов биологических специальностей.

Табл. 176. Ил. 42. Библиогр. 50 назв.

Рецензенты:

В. Ф. ЛЮБИМОВА, Н. А. ПЛОХИНСКИЙ, А. К. СКВОРЦОВ

ВВЕДЕНИЕ

Статистическая обработка цифрового материала, полученного путем опытов, учетов и наблюдений при биологических исследованиях, необходима для проверки степени достоверности результатов и правильного их обобщения. Особенно важна роль статистических методов, как средства, помогающего принять верное решение в условиях неопределенности. Каждый из методов математической статистики имеет свои возможности и ограниченную область применения. Использование метода, не соответствующего данному экспериментальному материалу, может привести к неверным обобщениям и необоснованным выводам.

Эта книга предназначена исследователям-биологам в качестве пособия для самостоятельного выбора метода математической статистики, наиболее соответствующего изучаемому вопросу, и для ознакомления с техникой расчетов при использовании избранного способа обработки цифровых данных.

Перед обработкой результатов исследования прежде всего необходимо определить задачи, которые должны быть решены математическим путем. Отдельные главы книги посвящены этим задачам, и они обычно следующие.

1. Вычисление средней арифметической и ее ошибки, коэффициента вариации и других показателей, характеризующих один ряд однородных чисел, полученных в эксперименте или иным путем (глава 1).

2. Установление по двум и больше сопряженным рядам чисел наличия связи (корреляции) между признаками (глава 2).

3. Определение по двум сопряженным рядам чисел формы зависимости (регрессии) одного признака от другого (глава 3).

4. Проверка гипотезы о различии (или сходстве) между признаками или вариантами опыта (глава 4).

Содержание первых четырех глав книги соответствует в той же последовательности четырем названным общим задачам. Иначе говоря, если в распоряжении исследователя одна совокупность, или группа чисел, то для ее математической обработки подойдут методы из главы 1; при двух и более различных группах чисел можно применить методы из 2, 3 и 4 глав.

В главе 5 излагаются специальные случаи применения математических методов в ботанике.

По оглавлению можно найти параграф, содержащий изложение метода, которым надлежит пользоваться. В начале каждого параграфа приводятся разъяснения, дающие возможность дальнейшего

уточнения метода обработки данных. Большинство методов биометрической обработки изложено в следующем порядке: а) случаи применения метода; б) ограничения его применения; в) формулы и объяснение символов; г) порядок расчетов по этапам; д) определение ошибки и достоверности полученных показателей.

В книге все таблицы обозначены числами без скобок, например, табл. 1.04 означает таблицу 4 в главе 1. Номера таблиц в приложении сопровождаются буквой П, например, 5П означает таблицу 5 в приложении. Формулы пронумерованы числами, заключенными в скобки, например, формула (2.13) означает формулу 13 в главе 2.

Пример подбора метода статистической обработки. Требуется проверить предположение о том, что между величиной зерен и продолжительностью их созревания у ячменя существует корреляционная взаимосвязь (табл. 2.11). В соответствии с задачей (см. выше) выбираем главу 2. В начале главы 2 читаем пояснения, затем составляем корреляционную решетку, так как численность выборки больше 30. По расположению частот корреляционной решетки (табл. 2.11) принимаем связь за криволинейную (она встречается значительно чаще прямолинейной, и во всех неясных случаях связь лучше считать криволинейной).

Любой из методов, указанных в § 2.06, 2.10, 2.11, 2.12, можно в рассматриваемом случае применить, но ввиду того, что деление наших данных на классы скорее качественное, чем количественное, выбираем § 2.12, следуя указанию в его начале.

Если требуется получить все остальные статистические показатели совокупности, то лучше сразу обратиться к объединенным схемам расчетов (см. § 1.16 и 1.22). Для получения предварительных выводов при оценке различий между небольшими совокупностями лучше воспользоваться непараметрическими критериями (см. § 4.09—4.13).

Цифровой материал, подлежащий обработке, должен быть предварительно просмотрен для выявления грубых ошибок, если таковые имеются. Числа, очень сильно отклоняющиеся от остальных, надо обязательно проверить на принадлежность их к данной совокупности (см. § 1.02). Все цифровые данные перед обработкой следует округлять до одинакового числа знаков после запятой, причем оно должно быть минимальным, но достаточным для достижения требуемой точности выводов.

Методы статистической обработки изложены нами в возможно лаконичной, почти рецептурной форме для того, чтобы яснее была видна схема расчетов. В целях сокращения вычислительных операций ряд формул был преобразован; предложены также некоторые новые формулы. Статистическая терминология объясняется в тех параграфах, в которых она впервые употребляется.

СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ВАРИАЦИОННОГО РЯДА

Все объекты каждого исследования (растения, животные, урожаи с опытных делянок или вегетационных сосудов, образцы плодов, семян и пр.) образуют так называемую общую, или генеральную, совокупность. Термин совокупность в статистике относят и к полученным в опыте или путем наблюдений числам, характеризующим с какой-либо одной количественной стороны объекты, входящие в данную генеральную совокупность. При этом в статистическую совокупность следует включать лишь числа, относящиеся к качественно однородным признакам (свойствам) объекта исследования. Значение отдельного члена совокупности называется вариантом, или датой. При большом числе дат в генеральной совокупности можно исследовать ее некоторую часть, которая называется выборкой (или частичной совокупностью). Достаточный объем выборки для изучения свойств генеральной совокупности определяется по методам из § 1.24.

§ 1.01. Составление вариационных рядов

Числа совокупности, или выборки (пример таких чисел дается в табл. 1.01), если последняя не очень велика по объему, желательно (но не обязательно) расположить в порядке убывания (или уве-

Таблица 1.01. Продолжительность вегетационного периода (в днях)
у 214 сортов ячменя

66	69	63	77	73	66	72	75	75	80	78	63	75	73	74	80	75	78
68	77	82	75	68	72	84	74	65	73	71	81	80	72	79	75	80	78
84	78	67	84	70	65	79	64	72	77	77	77	82	62	77	69	75	66
69	74	81	91	77	81	78	83	78	80	79	76	77	78	77	77	67	82
80	82	80	82	82	78	79	79	86	72	77	74	76	75	79	79	78	82
80	80	84	80	81	80	76	77	75	76	82	80	78	79	77	68	81	80
75	80	78	79	79	78	80	78	78	76	78	73	65	65	67	76	78	67
71	81	76	78	76	72	75	74	70	70	74	84	76	78	73	70	76	71
74	72	75	81	82	75	73	76	69	72	68	71	80	74	72	80	66	79
78	79	64	68	74	80	64	63	65	64	82	74	64	74	64	64	65	68
<u>60</u>	71	77	68	75	76	67	65	70	66	82	74	80	76	72	78	80	80
<u>69</u>	67	88	<u>92</u>	92	65	73	84	73	71	76	72	78	74	80	80		

Число вариант, или объем выборки $N = 214$.

Таблица 1.02. Высота растений сои (в см) в виде ранжированного невзвешенного вариационного ряда

Варианты
82
77
74
74
73
66
64
63
63
62
54
44
43
$\sum x = 839$
$N = 13$

Таблица 1.03. Данные таблицы 1.01 в виде ранжированного взвешенного вариационного ряда

Границы класса	Середина класса, x	Частота, f	Частость
57,5—62,4	60	2	0,9
62,5—67,4	65	30	14,0
67,5—72,4	70	34	15,9
72,5—77,4	75	62	29,0
77,5—82,4	80	74	34,6
82,5—87,4	85	8	3,7
87,5—92,5	90	4	1,9
$c = 5$		$k = 7$	$N = 214$
			100%

Таблица 1.04. Число мутантов среди растений ячменя

Класс	Число растений, или частота	Частость	
		%	доли единицы
1. Нормальные растения	984	93,71	0,937
2. Мутанты	66	6,29	0,063
Всего просмотрено растений, или объем выборки		$N = 1050$	100% 1,000

личения) их величин — в виде ранжированного невзвешенного вариационного ряда (например, табл. 1.02). Если совокупность, или выборка, содержит много чисел (например, табл. 1.01), то их следует сгруппировать в классы и составить ранжированный взвешенный вариационный ряд (табл. 1.03). Разновидностями таких вариационных рядов являются многоклассовый (табл. 1.03) и двухклассовый (табл. 1.04) ряды, причем последний часто называют рядом с альтернативной (дихотомической) изменчивостью вариантов.

Каждая группа чисел (вариант), на которые разбивается совокупность, или выборка, называется классом. Число классов зависит от величины классового интервала, т. е. разницы между наибольшим и наименьшим значениями в пределах каждого класса. Классовые интервалы для одного вариационного ряда обычно делают одинаковыми. Иногда делят совокупность на классы по качественным признакам, имеющим не строго определенное количественное содержание (например, ряды x и y в табл. 2.11).

Число вариантов, относящихся к данному классу, называется его частотой. Например, в табл. 1.03 класс 77,5—82,4 имеет частоту 74.

Частоты, выраженные в относительных числах, как доли единицы или проценты от общего числа вариантов (объема выборки), называют частостями (они приведены, например, в последних столбцах табл. 1.03 и 1.04).

Группировка вариантов по классам и составление таким образом взвешенных вариационных рядов представляет выборку в легко обозримом виде и облегчает последующие расчеты. Однако любая группировка, при которой истинные значения групп вариантов, объединяемых в классы, заменяются средними значениями классов, вносит неточность в получаемые затем показатели. Большей частью эта неточность невелика и ею можно пренебречь, однако в тех случаях, когда представляется к этому возможность, например, при обработке данных на электронно-вычислительной машине рекомендуется работать с невзвешенным рядом, без разбивки его на классы. Располагая лишь арифмометром и логарифмической линейкой (минимум вычислительных приборов для биометрических расчетов), целесообразно составлять взвешенный вариационный ряд при объеме совокупности свыше 50 вариантов, а при наличии настольного вычислительного автомата не менее 150 вариантов. Такие объемы данных еще доступны для обработки без их группировки, ибо последняя требует часто много времени и внимания. Составление взвешенных вариационных рядов поддается программированию и может быть выполнено на ЭВМ.

Взвешенный вариационный ряд составляют также и с целью определения типа распределения обрабатываемых данных. Распределение может быть нормальным (§ 1.25), биномиальным (§ 1.27), пуассоновским (§ 1.29) или иным. Если эмпирическое (полученное в данном опыте) распределение частот вариантов близко к нормальному (метод проверки этого соответствия указан в § 4.14), все приводимые ниже методы применимы к данному материалу (в необходимых случаях в тексте имеются указания об ограничениях применения метода в отношении того или иного типа распределения вариантов). Подавляющее большинство биологических признаков имеет симметричное распределение вариантов, близкое к нормальному. Согласно неравенству Чебышева из теории вероятностей, при любом типе распределения вариант вероятность отклонения последнего от основного закона нормального распределения, так называемого «правила трех сигм», составляет меньше, чем $\frac{1}{9}$ или равна этому числу. Следовательно, принимая неизвестное распределение вариантов за нормальное, можно ошибиться лишь в 11 случаях из 100. Составление взвешенных рядов необходимо также для построения шкалы балльной оценки (§ 5.01).

Порядок действий при составлении взвешенного вариационного ряда следующий.

1. Определяется величина классового интервала, которая зависит от принятого числа классов и от объема выборки:

$$c = \frac{x_N - x_1}{k - 1}, \quad (1.01)$$

где c — величина классового интервала; x_N — максимальная варианта выборки (в табл. 1.01 это 92); x_1 — минимальная варианта (в табл. 1.01 это 60); k — число классов, которое принимается обычно от 7 до 15 или по табл. 1П.

Для данных табл. 1.01 по формуле (1.01):

$$c = \frac{92 - 60}{7 - 1} \approx 5,$$

а число классов принято $k = 7$. Классовый интервал выражается целым и дробным числом в зависимости от того, являются ли варианты дискретными, т. е. неделимыми (особи, число зерен, дней и т. п.) или непрерывными, измеряемыми величинами (длина, ширина, высота, вес и т. п.).

Величину классового интервала можно рассчитать также по формуле:

$$c = \frac{(x_N - x_1) \lg 2}{\lg N}; \quad \lg 2 = 0,30102;$$

или:

$$c = \frac{(x_N - x_1) \ln 2}{\ln N} \quad \ln 2 = 0,69314; \quad (1.02)$$

или по формуле:

$$c = \frac{x_N - x_1}{\sqrt{N} - 1}, \quad (1.03)$$

где N — объем выборки, а прочие обозначения те же, что и в формуле (1.01).

По формуле (1.02) для данных табл. 1.01 классовый интервал равен:

$$c = \frac{(92 - 60) \lg 2}{\lg 214} \approx 4.$$

Такой интервал дал бы около 9 классов, однако для сокращения вычислений и устранения нулевых классов примем классовый интервал, подсчитанный выше, по формуле (1.01): $c = 5$ дней. О других формулах определения числа классов ряда см. § 5.01. Критерием правильности выбора величины классового интервала является четкое выявление основной тенденции распределения при наименьшей потере информации о форме изгибов кривой и других существенных особенностях распределения вариантов в данной совокупности. Иногда оказывается, что выборка состоит из разнокачественных элементов, и нулевые (а также очень малочисленные) классы служат границей между ними, что можно установить по виду двухвершинной кривой распределения частот и путем анализа собранного материала. Такие выборки для дальнейшей обработки обычно разделяют на две.

2. Расчет границ классов начинается от произвольного начала таким образом, чтобы полученные затем срединные значения классов были по возможности малозначными числами, что сократит объем

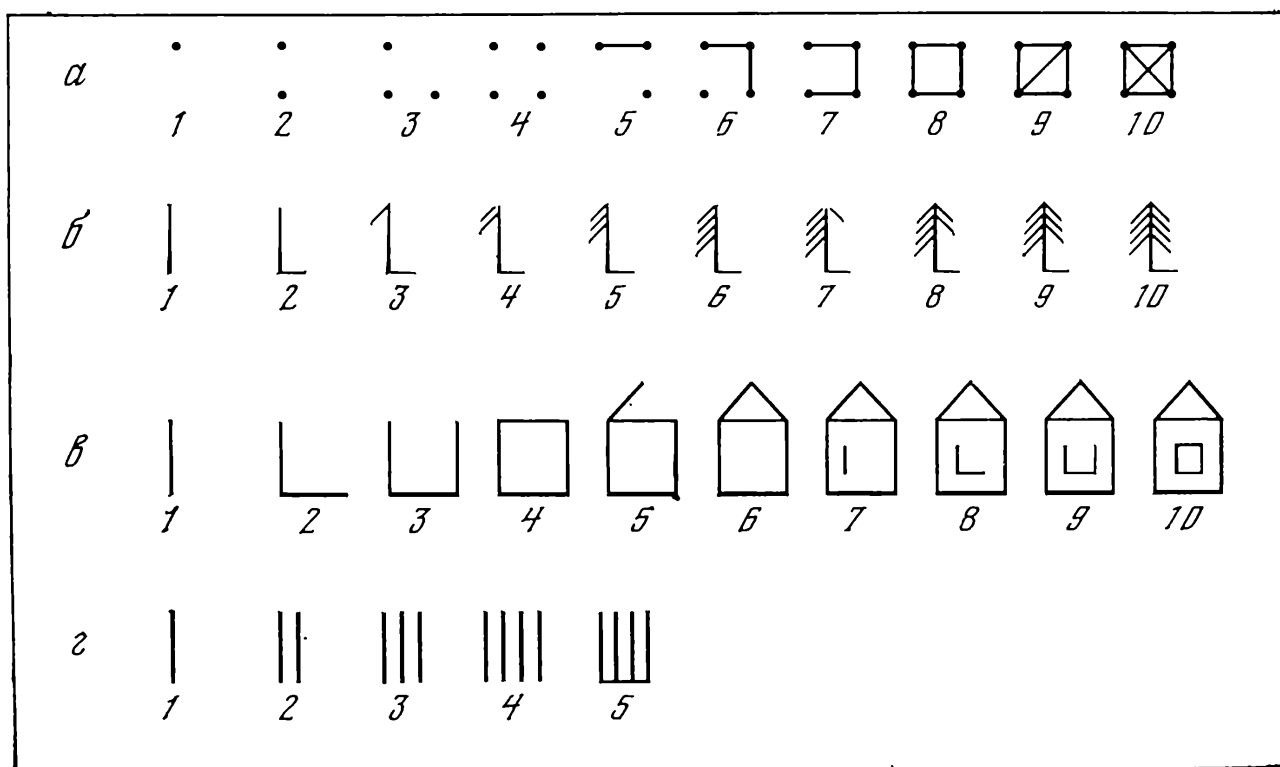


Рис. 1. Различные способы регистрации частот при составлении взвешенных вариационных рядов

вычислительной работы. Таким произвольным началом в табл. 1.03 является число 57,5, к которому прибавляем по 5 ($c = 5$), записывая все суммы (границы классов) вплоть до высшего класса. Другая граница первого класса — величина 62,4 — получена небольшим уменьшением значения верхней границы второго класса: $62,5 - 0,1 = 62,4$. Это целесообразно делать для удобства разности вариантов по классам. Все остальные нижние границы классов: 67,4, 72,4 и т. д. получаются путем последовательного прибавления к ним величины классowego интервала. Произвольное начало отсчета границ классов желательно выбирать так, чтобы крайние варианты x_1 и x_N по возможности оказались ближе к середине своих классовых интервалов.

3. Срединное значение первого класса в табл. 1.03 вычисляем так: $(57,5 + 62,5)/2 = 60$ дней. Все остальные срединные значения классов получаются путем последовательного прибавления к ним величины классowego интервала. Середины классов после их определения рассматриваются как варианты ряда и обозначаются также через x .

4. Разность вариантов по классам часто производят «конвертиком», т. е. точками и линиями, последовательные сочетания которых обозначают числа от 1 до 10 (рис. 1, а). Если регистрационные пометки при разности вариантов делать только штрихами, например знаками «елочка» и «домик», то это уменьшит вероятность ошибок при их подсчете (рис. 1, б, в). Обозначают число вариантов также пятерками штрихов (рис. 1, г). Сумма частот после разности должна равняться объему выборки (в табл. 1.03 $N = 214$), сумма процентных

частостей должна равняться 100%. Если совокупность вариантов не вызывает сомнений в отношении ее качественной однородности, но желательно устранить классы, в которых отсутствуют частоты; следует увеличить классовый интервал и повторить работу по разносте вариантов сначала. Когда нулевые классы находятся ближе к началу или концу ряда, можно отбросить крайние классы с их частотами (если число классов достаточно большое), благодаря чему не будут включены в ряд сильно отклоняющиеся варианты выборки, нередко возникающие из-за ошибок при сборе материала.

Если форма кривой распределения свидетельствует о «смешанном» распределении (две или несколько вершин), это желательно объяснить при биологическом толковании результатов.

§ 1.02. Оценка сильно отклоняющихся вариантов

Наибольшее значение признака в выборке называется максимальной вариантой, а наименьшее — минимальной вариантой. Если в ранжированном вариационном ряду значение максимальной или минимальной варианты чрезмерно велико или мало по сравнению с соседними числами ряда, то необходимо проверить принадлежность к данной совокупности таких резко отклоняющихся вариантов, которые могут относиться к иному сорту, другой популяции, а также возникнуть из-за грубой ошибки при измерении или подсчете. В подобных случаях резко отклоняющиеся варианты, чуждые данной совокупности, должны быть исключены из выборки. Однако это исключение должно быть обосновано посредством количественной оценки отклонения таких вариантов.

Для максимальной варианты (при нормальном или близком к нему распределении вариантов) применяется формула:

$$\vartheta_N = \frac{x_N - x_{N-1}}{x_N - x_2}, \quad (1.04)$$

где ϑ_N — критерий принадлежности максимальной варианты к совокупности; x_N — максимальная варианта; x_{N-1} — варианта, следующая по величине за максимальной; x_2 — варианта, стоящая в ранжированном ряду рядом с минимальной.

Проверим, например, принадлежность варианты 82 в табл. 1.02 к данной совокупности:

$$\vartheta_N = \frac{82 - 77}{82 - 44} = 0,132.$$

Так как вычисленное значение критерия, которое для $N = 13$ равно 0,520 (табл. 2П), меньше табличного, то варианту нельзя исключать из выборки.

Для минимальной варианты при распределении, не сильно отличающемся от нормального, применяется формула:

$$\vartheta_1 = \frac{x_2 - x_1}{x_{N-1} - x_1}, \quad (1.05)$$

где ϑ_1 — критерий принадлежности минимальной варианты x_1 к совокупности; x_2 — варианта, стоящая рядом с минимальной; x_{N-1} — вторая по значению варианта после максимальной.

Для примера оценим принадлежность варианты 43 в табл. 1.02 к данной совокупности:

$$\vartheta_1 = \frac{44 - 43}{77 - 43} = 0,029.$$

Полученное значение критерия меньше табличного (табл. 2П): $0,029 < 0,520$, поэтому отбрасывать варианту не следует.

В случае, если резко отклоняются по величине не только крайние варианты, но и соседние с ними, принадлежность их к данной совокупности оценивается по тем же формулам (1.04, 1.05). Например, из вариационного ряда 192, 187, 135, 127, 120, 111, 98, 87, 71, 66, 52, 47, 41, 12, 9 по формуле (1.04) надо проверить варианту 187, а по формуле (1.05) варианту 12. Если эти варианты окажутся чуждыми для данной выборки, то, естественно, варианты 192 и 9 также должны быть исключены. Если известно среднее квадратическое отклонение нормально распределенной совокупности, то оценку степени отклонения максимальной или минимальной варианты можно сделать более точно при помощи табл. 2аП и формул:

$$\vartheta_N = \frac{x_N - x_{N-1}}{\sigma}, \quad (1.06)$$

$$\vartheta_1 = \frac{x_2 - x_1}{\sigma}, \quad (1.07)$$

где σ — среднее квадратическое отклонение, остальные обозначения те же, что и в формулах (1.04, 1.05). В § 1.09 вычислена величина σ для данного ряда, которая равна 11,98.

Повторим процедуру оценки этих же вариантов по формулам (1.06, 1.07).

$$\vartheta_N = \frac{82 - 77}{11,98} = 0,42, \quad \vartheta_1 = \frac{44 - 43}{11,98} = 0,08.$$

По табл. 2аП критическое значение для объема совокупности $N = 10$ (принимая ближайшее меньшее число) при уровне значимости 0,05 (см. об уровнях значимости § 1.03) равно 1,46, что больше вычисленных значений. Следовательно, нет оснований отбрасывать как максимальную, так и минимальную варианты. Если варианта после описанных способов проверки отбрасывается, то вычисление σ следует повторить для новой уменьшенной выборки.

§ 1.03. Средняя арифметическая невзвешенного ряда. Уровни значимости

Средняя арифметическая, являясь основной характеристикой статистической совокупности, отражает уровень, по отношению к которому колеблются значения вариант в ней. Вычисление средней арифметической по данному методу производится для выборки, варианты которой не сгруппированы и представляют собой невзвешен-

ный ряд, например приведенный в табл. 1.02. При большом объеме выборки целесообразно составить взвешенный вариационный ряд и применить метод из § 1.04. Если при вычислении средней арифметической имеется в виду затем оценить достоверность ее значения и получить, кроме того, другие параметры совокупности, то лучше воспользоваться схемой расчета из § 1.16.

Ранжирование ряда при вычислении средних величин не обязательно. Формула вычисления средней арифметической:

$$M = \frac{\sum x}{N}, \quad (1.08)$$

где M — средняя арифметическая; $\sum x$ — сумма всех вариантов ряда; N — объем выборки.

Порядок работы (данные в табл. 1.02).

1. Подсчитывается число вариантов: $N = 13$.

2. Суммируются все варианты данной совокупности: $\sum x = 839$.

3. Сумма вариантов делится на число вариантов:

$$M = 839/13 = 64,5 \text{ см.}$$

Для оценки достоверности средней вычисляется ее ошибка по формуле:

$$m_M = \frac{\sigma}{\sqrt{N}}, \quad (1.09)$$

где m_M — ошибка средней арифметической; σ — среднее квадратическое отклонение, вычисление которого для данного ряда приведено в § 1.09, где оно равно 11,98; N — объем выборки.

Для ряда из табл. 1.02;

$$m_M = \frac{11,98}{\sqrt{13}} = \pm 3,3 \text{ см.}$$

Достоверность средней арифметической оценивают по формуле:

$$t = \frac{M}{m_M}, \quad (1.10)$$

где t — критерий Стьюдента; M — средняя арифметическая; m_M — ее ошибка.

Для рассматриваемого примера величина критерия равна: $t = 64,5/3,3 = 19,5$. Число степеней свободы, которое необходимо знать для определения достоверности средней арифметической при помощи табл. 3П, принимается равным: $\nu = N - 1 = 13 - 1 = 12$, где ν — число степеней свободы; N — объем выборки.

Числом степеней свободы считается число независимых отклонений отдельных вариантов от средней. Из отклонений независимыми являются все, кроме одного ($N - 1$) последнего, величина которого уже определена остальными отклонениями, и поэтому это отклонение не будет независимым.

Табличное значение критерия Стьюдента (табл. 3П) для 5% уровня значимости равно 2,179 и на 0,1% уровне значимости равно

4,318. Полученное выше значение критерия 19,5 значительно выше табличного, поэтому средняя арифметическая вполне достоверна даже при самой строгой оценке, т. е. на 0,1% уровне значимости.

Установлены следующие три уровня значимости в порядке возрастания строгости оценки достоверности биометрических показателей: $W_1 = 5\%$, $W_2 = 1\%$, $W_3 = 0,1\%$. Им соответствуют в том же порядке возрастания строгости оценки следующие доверительные уровни: $P_1 = 95\%$, $P_2 = 99\%$, $P_3 = 99,9\%$. Выраженные в долях единицы уровни значимости равны соответственно: $W'_1 = 0,05$; $W'_2 = 0,01$ и $W'_3 = 0,001$, а доверительные уровни в долях единицы представляются следующими числами: $P'_1 = 0,95$; $P'_2 = 0,99$; $P'_3 = 0,999$. Для биологических исследований во многих случаях достаточно принимать 5% уровень значимости, или 95% доверительный уровень (что одно и то же), при котором считают достаточным, если подтвердится существенность выводов в 95 случаях из 100.

Если вычисленное значение критерия Стьюдента будет меньше табличного (из табл. 3П), это может также означать, что недостаточен объем выборки, по которой найдена средняя арифметическая, или же колеблемость вариантов слишком велика, или данные неоднородны.

Средняя арифметическая в результатах обработки опытных данных должна приводиться вместе с ее ошибкой, например:

$$M = 64,5 \pm 3,3 \text{ см.}$$

Для проверки вычислений рекомендуется умножить среднюю арифметическую на объем выборки; полученное число должно равняться сумме вариантов. В данном примере:

$$MN = \Sigma x; \quad 64,538 \cdot 13 = 838,99,$$

что можно считать хорошим совпадением, так как фактическая сумма вариантов равна 839 (табл. 1.02). Сумма отклонений вариантов от средней арифметической невзвешенного ряда должна быть равна нулю, что также можно использовать для проверки вычислений:

$$\Sigma (x - M) = 0.$$

Для ряда из табл. 1.02 эта сумма отклонений равна 0,006 (табл. 1.10). Сумма отклонений вариантов от любого другого числа всегда будет больше суммы отклонений от средней арифметической. Прибавление ко всем вариантам ряда или вычитание от них одного и того же числа изменяет среднюю арифметическую на ту же величину.

Если объем общей, или генеральной, совокупности известен, то ошибка средней вычисляется по формуле:

$$m_M = \sigma \sqrt{\frac{N_0 - N}{N(N_0 - 1)}}, \quad (1.11)$$

где m_M — ошибка средней арифметической; σ — среднее квадратическое отклонение; N_0 — объем генеральной совокупности; N — объем выборки из нее.

В примере из § 1.04, где $\sigma = 6,07$, $N = 214$, предположим, что $N_0 = 900$ сортов, тогда:

$$m_M = 6,07 \sqrt{\frac{900 - 214}{214(900 - 1)}} = 0,36.$$

Ошибку средней арифметической вычисляют также через отклонения от средней арифметической по формуле:

$$m_M = \sqrt{\frac{\sum (x - M)^2}{N(N - 1)}}, \quad (1.12)$$

где m_M — ошибка средней арифметической; $\sum (x - M)^2$ — сумма квадратов отклонений вариантов от средней арифметической (она вычислена в табл. 1.10); N — объем выборки.

Для данных из табл. 1.02 по формуле (1.12):

$$m_M = \sqrt{\frac{1721,23}{13(13 - 1)}} = 3,3.$$

По формуле ошибки средней (1.09) можно найти также объем выборки, если он неизвестен, а σ и ошибка известны: $N = \left(\frac{\sigma}{m_M}\right)^2$

§ 1.04. Средняя арифметическая взвешенного ряда

Если кроме вычисления средней арифметической для большой группы наблюдений необходимо получить также среднее квадратическое отклонение, ошибку средней и другие параметры совокупности, то лучше пользоваться обобщенной схемой расчета из § 1.22.

а. Способ произведений.

Для больших выборок рекомендуется составить взвешенный вариационный ряд (см. § 1.01), средняя арифметическая затем вычисляется по формуле:

$$M = \frac{\sum xf}{\sum f} = \frac{\sum xf}{N}, \quad (1.13)$$

где M — средняя арифметическая; $\sum xf$ — сумма произведений середин классов x на соответствующие им частоты f ; N — объем выборки.

Вычисление M по данным о вегетационном периоде ячменя из табл. 1.03 показано в табл. 1.05. Расчеты проводятся в следующем порядке.

1. Умножаем варианты на их частоты $60 \cdot 2 = 120$ и т. д.

2. Суммируем эти произведения: $120 + 1950 + 680 + 360 = 16060$.

3. Полученную сумму $\sum xf = 16060$ делим на объем выборки: $M = 16060/214 = 75,05$.

4. Достоверность полученной средней арифметической оцениваем так же, как это было сделано в § 1.03. Среднее квадратическое отклонение для приводимых здесь данных о вегетационном периоде ячменя вычислено в § 1.10 и равно 6,07.

**Таблица 1.05. К вычислению средней арифметической
взвешенного ряда по способу произведений**

Границы класса	Середина класса (варианта), x	Частота, f	xf
57,5—62,4	60	2	120
62,5—67,4	65	30	1 950
67,5—72,4	70	34	2 380
72,5—77,4	75	62	4 650
77,5—82,4	80	74	5 920
82,5—87,4	85	8	680
87,5—92,5	90	4	360
$c = 5$	$k = 7$	$N = 214$	16 060

5. Получаем ошибку средней арифметической по формуле (1.09):

$$m_M = \frac{6,070}{\sqrt{214}} = \pm 0,415 \text{ дня.}$$

6. Вычисляем критерий Стьюдента по формуле (1.10): $t = 75,05 / 0,415 = 181$.

7. Число степеней свободы равно: $\nu = N - 1 = 214 - 1 = 213$.

8. По табл. 3П, на 5% уровне значимости при 213 степенях свободы $t = 1,960$. Полученное выше значение критерия достоверности (181) больше табличного, поэтому достоверность средней арифметической не вызывает сомнений. В окончательном виде средняя арифметическая для разбираемого примера записывается так: $M = 75,0 \pm 0,4$ дня.

Для проверки вычислений рекомендуется умножить объем выборки на среднюю арифметическую, что должно равняться сумме Σxf : $214 \cdot 75,05 = 16060,70$.

Полученная разница 0,70 может быть опущена, так как возникла вследствие принятой в данном расчете точности промежуточных вычислений.

Сумма $\Sigma (x - M) f$ у взвешенного ряда должна равняться нулю в пределах точности вычислений, что также можно использовать для проверки правильности вычислений средней арифметической.

б. Способ сумм.

В тех случаях, когда вариационный ряд состоит из большого числа классов (например, больше 12), а частоты и середины классов выражаются большей частью многозначными числами, вычисление M по способу произведений требует много времени из-за необходимости перемножения больших чисел. В подобных случаях и особенно при отсутствии достаточно производительной вычислительной техники можно применить способ суммирования частот вместо их умножения на варианты.

Вычислим способом сумм среднюю дату конца цветения гладиолуса в Москве (табл. 1.06). В таблице даты конца цветения в столбцах 1, 2 даются в числах от 1 марта (см. табл. 24П).

Таблица 1.06. Средняя арифметическая способом сумм

Класс, в днях от 1 марта	Середина класса, x	Частота, f	Ряд сумм частот	
			первый	второй
150—154,9	152,5	10	10	10
155—159,9	157,5	15	25	35
160—164,9	162,5	58	83	118
165—169,9	167,5	137	220	338
170—174,9	172,5	226	446	784
175—179,9	177,5	221	667	1451
180—184,9	182,5-A	230	—	—
185—189,9	187,5	145	458	1034
190—194,9	192,5	149	313	576
195—199,9	197,5	94	164	263
200—204,9	202,5	51	70	99
205—209,9	207,5	11	19	29
210—214,9	212,5	6	8	10
215—219,9	217,5	2	2	2
$c = 5$	$k = 14$	$N = 1355$		

В столбце 2 принимаем за условную среднюю — центральную варианту ряда: $A = 182,5$; при этом можно за условную среднюю взять любую варианту, но желательно ближе к середине ряда. Частоты из столбца 3 последовательно суммируем до черточки сверху вниз и в нижней половине ряда — снизу вверх, получаем первый ряд сумм частот, из которого таким же образом получаем второй ряд сумм частот. Проверку суммирования производим сложением трех центральных частот из столбцов 3 и 4: $230 + 667 + 458 = 1355$. Результат совпадает с объемом выборки; следовательно, нет ошибки в составлении первого ряда сумм частот. Средняя арифметическая вычисляется по формуле:

$$M = A + \frac{\delta_2 - \delta_1}{N} \cdot c, \quad (1.14)$$

где M — средняя арифметическая; A — условная средняя ряда, равная 182,5 в табл. 1.06; δ_2 — верхнее число нижней половины второго ряда сумм частот (равно 1034); δ_1 — нижнее число верхней половины второго ряда сумм частот (равно 1451); N — объем совокупности; c — классовой интервал.

По формуле (1.14):

$$M = 182,5 + \frac{1034 - 1451}{1355} \cdot 5 = 181,0.$$

Ошибка средней арифметической и ее достоверность определяются по формулам из 1.03.

§ 1.05. Общая средняя арифметическая по частным средним

Иногда встречается необходимость получить общую среднюю составной выборки, для каждой из отдельных частей которой средние арифметические уже вычислены. Такую общую среднюю можно найти по формуле:

$$M = \frac{\sum N_i M_i}{\sum N_i}, \quad (1.15)$$

где M — общая средняя арифметическая; N_i — объемы отдельных частей выборки; M_i — средние арифметические частей выборки; i — порядковые номера частей выборки; n — число частей выборки ($1 \leq i \leq n$).

Число отдельных частей выборки может быть любое. Если части выборки одинаковы по своему объему, т. е. $N_1 = N_2 = N_3 = \dots = N_n$, то формула (1.15) упрощается:

$$M = \frac{\sum M_i}{n}$$

В качестве примера найдем общую среднюю арифметическую высоты растений гелениума осеннего за четыре года наблюдений. Число растений и средняя их высота (в см) по годам были: $N_1 = 16$; $M_1 = 87$; $N_2 = 16$; $M_2 = 135$; $N_3 = 20$; $M_3 = 103$; $N_4 = 18$; $M_4 = 89$.

Общая средняя по формуле (1.15):

$$M = \frac{16 \cdot 87 + 16 \cdot 135 + 20 \cdot 103 + 18 \cdot 89}{16 + 16 + 20 + 18} = 103 \text{ см.}$$

Ошибка общей средней арифметической и ее достоверность определяются по формулам (1.09), (1.10), при этом за N принимается весь объем полной выборки, а среднее квадратическое отклонение рассчитывается по способу, указанному в § 1.12.

В указанном примере среднее квадратическое отклонение объединенной выборки равно 14,82, поэтому ошибка средней арифметической равна:

$$m_M = \frac{\sigma}{\sqrt{\sum N_i}} = \frac{14,82}{\sqrt{16 + 16 + 20 + 18}} = 1,77.$$

§ 1.06. Средняя гармоническая

При усреднении величин, представляющих собой изменения скорости каких-либо процессов, например прироста длины или диаметра побегов, а также при усреднении индексов, выражающих, например, число устьиц на 1 см^2 , или величин концентрации, например 1 мг на 1 л , или других подобных отношений, вместо средней арифметической пользуются средней гармонической, которая для

невзвешенного вариационного ряда вычисляется по формуле:

$$H = \frac{N}{\sum \frac{1}{x}}, \quad (1.16)$$

где H — средняя гармоническая; N — объем выборки; x — варианты выборки.

Пример. В результате подсчетов числа устьиц на определенной площади по пяти образцам листьев березы бородавчатой получены индексы, представляющие собой число устьиц на 1 см^2 (табл. 1.07), которые оказались равными 5, 7, 10, 13, 15 устьиц на 1 см^2 . Средняя арифметическая этих индексов равна:

$$M = \frac{5 + 7 + 10 + 13 + 15}{5} = \frac{50}{5} = 10 \text{ устьиц/см}^2.$$

Средняя гармоническая этих же чисел по формуле (1.16) равна:

$$H = \frac{5}{\frac{1}{5} + \frac{1}{7} + \frac{1}{10} + \frac{1}{13} + \frac{1}{15}} = \frac{5}{0,587} = 8,52 \text{ устьиц/см}^2.$$

Прямым подсчетом можно доказать, что в данном случае более правильно употреблять H вместо средней арифметической. Так, общая площадь пяти листьев (по одному от каждого образца) при условии, что на каждом из них будет 100 устьиц, равна:

$$\frac{100}{5} + \frac{100}{7} + \frac{100}{10} + \frac{100}{13} + \frac{100}{15} = 58,7 \text{ см}^2.$$

Общее число устьиц на этой поверхности, определенное через H , равно: $58,7 \cdot 8,52 = 500,1$ устьиц, в то время как число устьиц на той же поверхности, определенное через $M = 10$, составляет $58,7 \cdot 10 = 587$ устьиц. Между тем, по условию задачи взято пять листьев по 100 устьиц на каждом, т. е. всего должно быть именно $5 \cdot 100 = 500$ устьиц. Как видим, средняя гармоническая правильно отражает суть дела, а средней арифметической пользоваться здесь нельзя.

§ 1.07. Средняя квадратическая и средняя кубическая

При выведении средних величин для значений площади и объема необходимо согласовать средние по линейным размерам и средние получаемых по ним значений площади и объема. Вычисленная по линейным размерам средняя арифметическая обычно дает меньшие значения площади и объема, чем те, которые получают по фактическому линейному размеру. В таких случаях вычисляют среднюю квадратическую для значений линейного параметра поверхности и среднюю кубическую для значений линейного параметра объема.

Средняя квадратическая невзвешенного вариационного ряда вычисляется по формуле:

$$S = \sqrt{\frac{\sum x^2}{N}}, \quad (1.17)$$

Таблица 1.07. К вычислению средней гармонической

Номер образца	Число устьиц на см ² , x	$\frac{1}{x}$	$\frac{100}{x}$
1	5	0,200	20,0
2	7	0,143	14,3
3	10	0,100	10,0
4	13	0,077	7,7
5	15	0,067	6,7
	50	0,587	58,7 см ²

Таблица 1.08. Размеры листьев примулы

Длина листовой пластинки, см, x	Фактическая площадь листа, см ² , y
3	7
5	20
6	30
8	55
$\sum x = 22$ $M_x = 5,500$ $S = 5,788$	$\sum y = 112$ $M_y = 28,000$

где S — средняя квадратическая; x — отдельные варианты линейного параметра; $\sum x^2$ — сумма квадратов вариантов; N — объем выборки.

Форма участков поверхности может быть в принципе любой, однако одинаковой или близкой в пределах усредняемой выборки, значения площади в которой должны определяться по одной формуле. Например, по данным табл. 1.08 вычислим среднюю квадратическую длины листовой пластинки примулы обратноконической и покажем, что пользоваться средней арифметической в этом случае было бы неверно. Площадь листьев примулы с хорошей степенью совпадения может быть вычислена по уравнению степенной функции:

$$y = 0,6728 x^{2,1229},$$

где y — площадь листа; x — длина листовой пластинки.

Средние арифметические ряда x и ряда y равны:

$$M_x = 5,500 \text{ см и } M_y = 28,000 \text{ см}^2.$$

Если же подсчитаем среднюю площадь листа (y) по уравнению, подставив в него среднюю арифметическую длины листовой пластинки $x = 5,500$, то получим:

$$y = 0,6728 \cdot 5,500^{2,1229} = 25,000 \text{ см}^2,$$

т. е. средняя арифметическая длины листовой пластинки не соответствует истинной величине. Средняя квадратическая длины листовой пластинки:

$$S = \sqrt{\frac{3^2 + 5^2 + 6^2 + 8^2}{4}} = 5,7879.$$

Определяя среднюю площадь листа через среднюю квадратическую, получим:

$$y = 0,6728 \cdot 5,7879^{2,1229} = 28,000 \text{ см}^2,$$

т. е. то же значение, которое получается по непосредственным значениям площади. Следовательно, средняя квадратическая более верно передает зависимость площади от линейного параметра.

Средняя квадратическая взвешенного ряда вычисляется по формуле:

$$S = \sqrt{\frac{\sum fx^2}{N}}, \quad (1.18)$$

где S — средняя квадратическая; $\sum fx^2$ — сумма произведений частот (f) на квадраты соответствующих вариантов (x); N — объем выборки.

Средняя кубическая вычисляется по формуле (1.19) для определения средней величины линейного параметра, на основании которого рассчитывается объем какого-либо тела (шаровидных плодов, пыльцы цветков и т. д.):

$$Q = \sqrt[3]{\frac{\sum x^3}{N}}, \quad (1.19)$$

где Q — средняя кубическая; x — отдельные варианты линейного параметра; $\sum x^3$ — сумма их кубов; N — объем выборки.

Например, требуется определить средний радиус трех шаров, имеющих следующие радиусы и объемы: 5 см и 522,0250 см³, 6 см и 902,0592 см³, 7 см и 1432,4366 см³. Средние арифметические радиуса и объема соответственно 6 см и 952,1736 см³. Однако, если определять средний объем через среднюю арифметическую радиуса, равную 6, то получим средний объем 902,0592, т. е. величину, не совпадающую с фактической. Средняя кубическая значений радиусов:

$$Q = \sqrt[3]{\frac{5^3 + 6^3 + 7^3}{3}} = 6,1091 \text{ см.}$$

Средний объем шара, вычисленный по формуле с радиусом, равным средней кубической — 6,1091, в точности совпадает с фактическим средним объемом — 952,1736 см³.

Среднюю кубическую для взвешенного вариационного ряда вычисляют по формуле:

$$Q = \sqrt[3]{\frac{\sum fx^3}{N}}, \quad (1.20)$$

где $\sum fx^3$ — сумма произведений частот на соответствующие им кубы вариантов; N — объем выборки.

Проверка правильности вычисления средней квадратической и средней кубической линейных параметров заключается в том, что полученные средние при подстановке в формулы определения площади или объема должны давать фактические площадь или объем.

§ 1.08. Средняя геометрическая

При математической обработке данных по темпам количественного изменения явлений, например при усреднении темпов роста растений, вместо средней арифметической применяют среднюю геометрическую, которая вычисляется по формуле:

$$G = \sqrt[N]{\prod x}, \quad (1.21)$$

где G — средняя геометрическая; \prod — знак произведения, означает, что все числа, стоящие за ним, следует перемножить, т. е. $\prod x = x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_N$, среди них при этом не должно быть ни одного $x_i = 0$; N — число вариантов (например, число периодов времени и др.).

При вычислении средней геометрической обычно пользуются логарифмами:

$$\lg G = \frac{\sum \lg x}{N},$$

где $\sum \lg x$ — сумма логарифмов всех вариантов; N — объем выборки.

Темпы роста площади листьев (табл. 1.09) определяются делением площади последующего наблюдения на площадь предыдущего замера:

$$2,2 : 1,3 = 1,69; \quad 3,1 : 2,2 = 1,41 \text{ и т. д.}$$

Средняя арифметическая из значений темпов роста:

$$M = \frac{1,69 + 1,41 + 1,19 + 1,24}{4} = 1,38,$$

а средняя геометрическая:

$$\lg G = \frac{0,2279 + 0,1492 + 0,0755 + 0,0934}{4} = 0,13652,$$

отсюда $G = 1,37$.

Проверка ее вычисления производится по формуле:

$$x_N = x_1 G^{N-1} \quad (1.22)$$

где x_N — максимальная варианта, т. е. последнее измерение наблюдавшегося признака; x_1 — начальное измерение признака; G — средняя геометрическая; N — число наблюдений. $x_N = 1,3 \cdot 1,37^4 = 4,58$.

Полученное по формуле (1.22) значение максимальной варианты достаточно близко к фактическому: 4,6 см²; следовательно, средняя геометрическая вычислена верно. Подставляя в формулу (1.22) значение $M = 1,38$ вместо $G = 1,37$, получаем $x_N = 4,71$, т. е.

Таблица 1.09. Рост листа
зимолости

Сроки измерения, день	Средняя площадь листа, см ² , x	Темп роста, $\frac{x_i}{x_{i-1}}$
10-й	1,3	—
20-й	2,2	1,69
30-й	3,1	1,41
40-й	3,7	1,19
50-й	4,6	1,24

больше фактического: $4,71 > 4,6$. Поэтому средней арифметической здесь пользоваться было бы неправильно.

Средняя геометрическая взвешенного ряда вычисляется по формуле:

$$G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \dots x_k^{f_k}}, \quad (1.23)$$

или

$$\lg G = \frac{1}{N} (f_1 \lg x_1 + f_2 \lg x_2 + f_3 \lg x_3 + \dots + f_k \lg x_k),$$

где G — средняя геометрическая; x — варианты (середины классов); f — соответствующие им частоты; N — объем выборки, $N = \sum f$; k — число классов.

Между рассмотренными выше средними значениями одной и той же совокупности существует следующее соотношение: $H \leq G \leq M \leq S \leq Q$, т. е. соответственно по величине средние располагаются от меньшей к большей в следующем порядке: гармоническая, геометрическая, арифметическая, квадратическая, кубическая. Степень различия между средними одной и той же совокупности зависит от величины коэффициента вариации данных этой совокупности: чем больше коэффициент вариации, тем больше различаются по величине указанные средние. (О коэффициенте вариации см. § 1.13.)

Все средние величины на графике кривой распределения вариантов представляют собой точки на оси абсцисс.

§ 1.09. Среднее квадратическое отклонение невзвешенного ряда

О невзвешенном ряде см. § 1.01.

Мера варьирования признака в статистической совокупности имеет важное методологическое значение в биометрическом исследовании для углубленного познания изучаемого явления. Величина колебаний значений вариантов около их средней арифметической измеряется средним квадратическим отклонением. Вместе со средней арифметической среднее квадратическое отклонение (σ), или сигма, как его часто называют, является основой для вычисления многих параметров выборки и других биометрических показателей. Среднее квадратическое отклонение в квадрате называется дисперсией (σ^2). Суть понятия «среднее квадратическое отклонение» видно из табл. 1.10, где отклонения от средней перед их усреднением возведены в квадрат, благодаря чему все отклонения становятся положительными, а также получают большую величину, что повышает их значимость и, следовательно, определяет большую чуткость сигмы при измерении ею варьирования признака. Формула (1.24) чаще употребляется при работе на ЭВМ; при ручных расчетах, для невзвешенных рядов обычно применяют формулу (1.25).

а. Прямой способ.

Вычисление сигмы непосредственно через квадраты отклонений от средней требует меньшего числа операций, чем по остальным

способам:

$$\sigma = + \sqrt{\frac{\sum (x - M)^2}{N - 1}}, \quad (1.24)$$

где σ — среднее квадратическое отклонение; Σ — знак суммирования; x — варианты совокупности; M — средняя арифметическая; N — объем выборки.

Перед знаком корня в формуле (1.24) всегда принимается знак плюс.

Вычисление сигмы для невзвешенного ряда по этой формуле проводится в следующем порядке (табл. 1.10).

1. Определяется средняя арифметическая. Для данного ряда она вычислена в § 1.03 и равна $M = 64,538$.

2. Находится отклонение вариантов путем вычитания от каждой из них средней арифметической: $x - M$. Вычитая: $82 - 64,538 = 17,462$; $77 - 64,538$ и т. д., получим столбец 2 табл. 1.10, сумма которого должна равняться нулю. В нашем случае: $\Sigma (x - M) = 0,006$, что можно считать удовлетворительным приближением к нулю.

3. Возводятся в квадрат отклонения и получается их сумма:

$$\Sigma (x - M)^2 = 1721,230772.$$

4. Вычисляется сигма по формуле (1.24):

$$\sigma = + \sqrt{\frac{1721,230772}{13 - 1}} = 11,98.$$

В промежуточных вычислениях показателей обычно сохраняется число знаков, достаточное для получения необходимой его точности, сам же показатель, в данном случае сигма, приводится в результате с числом знаков, имеющим реальное значение.

б. Способ условной средней.

Для облегчения расчетов по данному способу, особенно при многозначных вариантах, используется то свойство среднего квадратического отклонения, что его величина не изменяется в зависимости от величины числа, от которого отсчитываются отклонения. Если прибавить ко всем вариантам ряда одно и то же число или вычесть его из них, то величина сигмы от этого не изменится. Вместо средней арифметической можно взять любое число A . Если взять это число целым, близким к средней арифметической, или равным нулю, то вычисления становятся менее трудоемкими. Для вычисления сигмы невзвешенного ряда при ручном счете наиболее часто используются формулы (1.25) и (1.27).

Вычислим сигму невзвешенного вариационного ряда через отклонения вариантов от условной средней $A = 64$. Для этого понадобятся только столбцы 1, 3, 4 табл. 1.11. В столбце 3 таблицы приведены разности числа A с каждой из вариантов. Сумма разностей с учетом знаков равна $+7$. В столбце 4 разности возведены в квадрат, сумма квадратов разностей равна 1725. Проверка промежуточных вычислений делается повторением расчета с другой условной сред-

Таблица 1.10. Вычисление суммы квадратов отклонений по данным табл. 1.02

Варианта, x	$x - M$	$(x - M)^2$
82	17,462	304,921444
77	12,462	155,301444
74	9,462	89,529444
74	9,462	89,529444
73	8,462	71,605444
66	1,462	2,137444
64	-0,538	0,289444
63	-1,538	2,365444
63	-1,538	2,365444
62	-2,538	6,441444
54	-10,538	111,049444
44	-20,538	421,809444
43	-21,538	463,885444
0,006		1721,230772

Таблица 1.11. Вычисление суммы квадратов отклонений от условной средней A и от нуля

Варианта, x	Квадрат варианты, x^2	$x - A$	$(x - A)^2$	$(x - 66)$
82	6 724	18	324	16
77	5 929	13	169	11
74	5 476	10	100	8
74	5 476	10	100	8
73	5 329	9	81	7
66	4 356	2	4	0
$A = 64$	4 096	0	0	-2
63	3 969	-1	1	-3
63	3 969	-1	1	-3
62	3 844	-2	4	-4
54	2 916	-10	100	-12
44	1 936	-20	400	-22
43	1 849	-21	441	-23
839	55 869	+7	1725	-19

ней, при этом должно выполняться условие:

$$N |A_1 - A_2| = |\Sigma (x - A_1)| + |\Sigma (x - A_2)|,$$

где N — объем совокупности; A_1, A_2 — первая и вторая условные средние; $\Sigma (x - A_1)$ и $\Sigma (x - A_2)$ — суммы отклонений вариантов от этих средних.

В столбце 5 табл. 1.11 приведены отклонения от второй условной средней $A_2 = 66$, вычисленные только с целью проверки промежуточных расчетов. Три члена последней формулы заключены в прямые скобки. Это обозначает, что перед ними в любом случае ставится знак плюс. По нашим данным: $13 |64 - 66| = |7| + |19|$; $26 = 26$. Следовательно, соответствующие расчеты выполнены верно. Знак суммы отклонений зависит от положения условной средней: если $A > M$, то $\Sigma (x - A) < 0$, а если $A < M$, то $\Sigma (x - A) > 0$.

Сигму вычисляем по формуле:

$$\sigma = + \sqrt{\frac{\Sigma (x - A)^2 - [\Sigma (x - A)]^2 / N}{N - 1}}, \quad (1.25)$$

где σ — среднее квадратическое отклонение; $\Sigma (x - A)^2$ — сумма квадратов отклонений вариантов от условной средней $A = 64$; $[\Sigma (x - A)]^2$ — сумма отклонений, возведенная в квадрат; N — число вариантов.

По формуле (1.25) получаем:

$$\sigma = + \sqrt{\frac{1725 - 7^2/13}{13 - 1}} = 11,98.$$

Средняя арифметическая может быть при этом получена по формуле:

$$M = A + \frac{\sum (x - A)}{N}, \quad M = 64 + \frac{+7}{13} = 64,54. \quad (1.26)$$

Обозначения в формуле (1.26) те же, что и в формуле (1.25).

Когда условные отклонения отсчитываются не от числа, близкого к средней арифметической, а от нуля ($A = 0$), тогда формула (1.25) примет вид:

$$\sigma = + \sqrt{\frac{\sum x^2 - (\sum x)^2/N}{N - 1}}, \quad (1.27)$$

где σ — среднее квадратическое отклонение; $\sum x^2$ — сумма квадратов вариант; $(\sum x)^2$ — квадрат суммы вариант; N — объем выборки.

Формулу (1.27) рекомендуется применять при однозначных вариантах. Если данные двузначные и выше и тем более если ряды большие, то рекомендуется формула (1.25), которая, значительно облегчая вычисления, не делает их менее точными.

Для вычисления сигмы по формуле (1.27) нужны лишь столбцы 1 и 2 табл. 1.11.

1. Суммируем варианты: $\sum x = 839$.

2. Возводим в квадрат все варианты и суммируем их: $\sum x^2 = 55869$.

3. По формуле (1.27) вычисляем:

$$\sigma = \sqrt{\frac{55869 - 839^2/13}{13 - 1}} = + 11,98.$$

в. Способ квадрата средней арифметической.

Наиболее просто сигма вычисляется по способу условной средней. Приведенная ниже формула ее расчета может употребляться в тех случаях, когда уже известна с достаточной точностью средняя арифметическая данного ряда:

$$\sigma = + \sqrt{\left(\frac{\sum x^2}{N} - M^2 \right) \frac{N}{N - 1}}, \quad (1.28)$$

где σ — среднее квадратическое отклонение; $\sum x^2$ — сумма квадратов вариант; N — объем выборки; M — средняя арифметическая данного ряда. При больших объемах выборок множитель $N/(N - 1)$ можно принять за единицу.

Для данных табл. 1.11 по формуле (1.28):

$$\sigma = \sqrt{\left(\frac{55869}{13} - 64,538^2 \right) \frac{13}{13 - 1}} = 11,98.$$

Определение ошибки сигмы и метод проверки ее вычисления см. § 1.10.

§ 1.10. Среднее квадратическое отклонение взвешенного ряда

а. Способ сумм.

Вычисление сигмы по способу сумм наиболее просто, однако в случае резко асимметричного ряда требуется подбор подходящего способа (§ 1.18). Данный способ сумм целесообразно применять при ручном счете и симметричных рядах, если варианты и частоты ряда — большие числа, особенно если ряд сравнительно длинный, например при числе классов больше 12. Вычислим сигму по этому способу для вариационного ряда распределения сортов ячменя по продолжительности вегетационного периода в днях, которая приведена в столбце 1 табл. 1.12 (середины классов), а в столбце 2 — соответствующее число сортов ячменя, или частоты.

1. Требуемый для расчетов вариационный ряд в нашем примере уже составлен (в табл. 1.12 приведены данные из табл. 1.03).

2. Против среднего класса ряда (его варианта — 75), в столбцах 3 и 4 ставим четыре черточки точно так, как это показано в табл. 1.12.

3. Против крайних частот (4 и 2) записываем те же самые частоты в столбцах 3 и 4.

4. В столбце 3 составляем первый ряд накопленных частот сверху вниз: $4 + 8 = 12$; $12 + 74 = 86$. Поскольку дальше вниз стоит черта, накопление частот прекращаем. Таким же образом, но снизу вверх получаем накопленные частоты нижней половины первого ряда: $2 + 30 = 32$; $32 + 34 = 66$; дальше идет черта, накопление прекращаем.

5. Составляем второй ряд накопленных частот сверху: $4 + 12 = 16$, ниже идут три черты. Частоты второго ряда, начиная снизу, накапливаем так: $2 + 32 = 34$, выше идут те же три черты, накопление частот закончено. Проверка правильности составления обоих рядов накопленных частот заключается в том, что три центральных числа вокруг первой черты должны в сумме дать объем выборки: $62 + 86 + 66 = 214 = N$, а прочие числа около четырех черточек равны: $86 + 16 = 102 = q_1$; $66 + 34 = 100 = q_3$.

Таблица 1.12. Вычисление полусумм накопленных частот

Варианта,	Частота, f	Первый ряд накопленных частот	Второй ряд накопленных частот
		$q_1 = 102$	$q_2 = 20$
90	4	4	4
85	8	12	16
80	74	86	—
75	62	—	—
70	34	66	—
65	30	32	34
60	2	2	2
$c = 5$	$N = 214$	$q_3 = 100$	$q_4 = 36$

6. Суммируем верхнюю и нижнюю половины каждого из двух рядов накопленных частот:

$$\begin{aligned} q_1 &= 4 + 12 + 86 = 102; & q_2 &= 4 + 16 = 20; \\ q_3 &= 66 + 32 + 2 = 100; & q_4 &= 34 + 2 = 36. \end{aligned}$$

7. Вычисляем среднее квадратическое отклонение по формуле:

$$\sigma = c \sqrt{\frac{q_1 + q_3 + 2(q_2 + q_4) - (q_1 - q_3)^2/N}{N - 1}}, \quad (1.29)$$

где σ — среднее квадратическое отклонение; c — классовой интервал; q_1 — верхняя полусумма первого ряда накопленных частот; q_3 — нижняя полусумма первого ряда; q_2 — верхняя полусумма второго ряда накопленных частот; q_4 — нижняя полусумма второго ряда; N — объем выборки.

Средняя арифметическая в данном случае попутно вычисляется по формуле:

$$M = A + \frac{c(q_1 - q_3)}{N}, \quad (1.30)$$

где M — средняя арифметическая; A — начало отсчета полусумм; c — классовой интервал; q_1 — верхняя полусумма первого ряда накопленных частот; q_3 — нижняя полусумма того же ряда; N — объем выборки.

Подставляя в формулу (1.30) данные табл. 1.12, получаем:

$$M = 75 + \frac{5(102 - 100)}{214} = 75,05.$$

Подставляя значения полусумм из табл. 1.12 в формулу (1.29), получаем:

$$\sigma = 5 \sqrt{\frac{102 + 100 + 2(20 + 36) - (102 - 100)^2/214}{214 - 1}} = 6,06.$$

Ошибка среднего квадратического отклонения вычисляется по формуле:

$$m_\sigma = \frac{\sigma}{\sqrt{2N}}, \quad (1.31)$$

или (при небольших выборках):

$$m_\sigma = \frac{\sigma}{\sqrt{2(N-1)}}, \quad (1.32)$$

где m_σ — ошибка сигмы; σ — среднее квадратическое отклонение; N — объем выборки.

Для сигмы, полученной по данным табл. 1.12, по формуле (1.31) ошибка равна:

$$m_\sigma = \frac{6,06}{\sqrt{2 \cdot 214}} = \pm 0,29.$$

Таким образом, в нашем примере $\sigma = 6,06 \pm 0,29$.

Если имеется предположение, что распределение частот исследуемого ряда сильно отличается от нормального, то лучше вычислять ошибку сигмы по более универсальной формуле:

$$m_{\sigma} = \sqrt{\frac{\mu_4 - \mu_2^2}{4\mu_2 N}}, \quad (1.33)$$

где m_{σ} — ошибка среднего квадратического отклонения; μ_4 — центральный момент четвертого порядка (см. § 1.18); μ_2 — центральный момент второго порядка (см. § 1.18); N — объем выборки.

Подставляя значения моментов, вычисленные в § 1.18, для той же сигмы получим ошибку по формуле (1.33):

$$m_{\sigma} = \sqrt{\frac{3519,947 - 36,68^2}{4 \cdot 36,68 \cdot 214}} = 0,263.$$

Несколько сокращает вычисления формула:

$$m_{\sigma} = 0,5 \sqrt{\left(\frac{\mu_4}{\mu_2} - \mu_2\right) \frac{1}{N}}. \quad (1.34)$$

При известной ошибке средней арифметической (m_M) можно получить тот же результат по формуле:

$$m_{\sigma} = 0,5 \sqrt{\frac{\mu_4}{\mu_2 N} - m_M^2}, \quad (1.35)$$

где обозначения те же, что и в формуле (1.33).

Приблизительная проверка вычисления сигмы может быть сделана по соотношениям:

$$x_N \approx M + 3\sigma; \quad x_1 \approx M - 3\sigma,$$

где x_N — максимальная варианта выборки; M — средняя арифметическая; σ — среднее квадратическое отклонение; x_1 — минимальная варианта.

Для рассматриваемого здесь примера:

$$x_N = 75,047 + 3 \cdot 6,06 \approx 93 \text{ дня}; \quad x_1 = 75,047 - 3 \cdot 6,06 \approx 57 \text{ дней}.$$

Фактически: $x_N = 92$ дня, а $x_1 = 60$ дней (см. табл. 1.01), что можно считать удовлетворительным совпадением.

Вычисление ошибок среднего квадратического отклонения и средней арифметической можно приближенно проверить при помощи существующего соотношения между ними:

$$\frac{m_{\sigma}}{m_M} = 0,70711.$$

Для рассматриваемого примера: $0,29/0,415 = 0,7$, что можно считать удовлетворительным результатом.

Таблица 1.13. К расчету сигмы взвешенного ряда по способу отклонений от условной средней

Варианта (середина класса), x	Частота, f	Условное отклонение, a	fa	fa^2		fa
90	4	+3	+12	36	2	8
85	8	+2	+16	32	1	8
80	74	+1	+74	74	0	0
$A=75$	62	0	0	0	-1	-62
70	34	-1	-34	34	-2	-68
65	30	-2	-60	120	-3	-90
60	2	-3	-6	18	-4	-8
<hr/>						
$c=5$	$N=214$	$a = \frac{x-A}{c}$	+2	314	$A=80$	-212

6. Способ условной средней.

Ниже приведенная схема расчета сигмы взвешенного ряда является наиболее распространенной и применяется для сокращения вычислительной работы, особенно тогда, когда варианты (середины классов) и соответствующие им частоты выражены многозначными числами. Расчет необходимых сумм для подстановки в формулу (1.36) приведен в табл. 1.13, где обработан ряд из табл. 1.03.

$$\sigma = +c \sqrt{\frac{\sum fa^2 - (\sum fa)^2/N}{N-1}} \quad (1.36)$$

где σ — среднее квадратическое отклонение; c — классовой интервал; f — частоты ряда; a — условные отклонения вариантов от $A = 75$; N — объем выборки.

Порядок действий.

1. Выбираем начальную точку отсчета (условную среднюю), расположенную в середине ряда и возможно ближе к максимальной частоте: $A = 75$.

2. Против класса с вариантой 75 (в столбце 3) записываем условное отклонение 0, затем вверх с плюсом, а вниз с минусом записываем отклонения в виде порядковых чисел: 1, 2, 3, 4 и т. д. по числу классов. Если варианты ряда (столбец 1) возрастают не так, как в табл. 1.13, а наоборот, т. е. сверху вниз, то отклонения из верхней половины ряда будут отрицательными, а из нижней половины ряда — положительными.

3. Перемножаем условные отклонения на частоты f : $+3 \cdot 4 = +12$; $+2 \cdot 8 = +16$ и т. д. (столбец 4).

4. Числа столбца 4 умножаем по строкам на соответствующие числа столбца 3: $12 \cdot 3 = 36$; $16 \cdot 2 = 32$ и т. д. (столбец 5).

5. Суммируя числа столбцов 4 и 5, получаем:

$$\sum fa = +2 \text{ и } \sum fa^2 = 314.$$

6. Подставляя суммы +2 и 314 в формулу (1.36), получим:

$$\sigma = 5 \sqrt{\frac{314 - 2^2/214}{214 - 1}} = 6,06.$$

Среднюю арифметическую получаем по формуле:

$$M = A + \frac{c \sum fa}{N}, \quad (1.37)$$

где A — начало отсчета условных отклонений, или условная средняя; c — классовой интервал; M — средняя арифметическая; $\sum fa$ — сумма произведений частот на условные отклонения; N — объем выборки. Частное $(\sum fa)/N$ называется первым условным моментом m_1 (см. § 1.18 и 1.22).

В приведенном примере:

$$M = 75 + \frac{5 \cdot 2}{214} = 75,05.$$

7. Проверку вычисления M делаем повторением вычислений суммы отклонений от другой условной средней. В столбцах 6 и 7 исключительно для этой цели составлены отклонения от новой условной средней $A_2 = 80$ и произведения отклонений на те же частоты. Если $A_1 < M < A_2$, то должно выполняться условие: $|\sum_1 fa| + |\sum_2 fa| = N$, если $A_1 < A_2 < M$ или $A_1 > A_2 > M$, то должно выполняться условие: $|\sum_1 fa| - |\sum_2 fa| = N$. Заключение сумм в прямые скобки означает, что независимо от их фактического знака, перед ними всегда ставится знак плюс. Так как $75 < M = 75,05 < 80$, то выбираем первую из двух формул проверки и получаем: $2 + 212 = 214$, т. е. условие выполняется. Средняя арифметическая, вычисленная по второй сумме отклонений $\sum_2 fa = -212$, равна:

$$M = 80 + \frac{5 \cdot (-212)}{214} = 75,05,$$

т. е. тому же значению, что и при вычислении по первой сумме отклонений. Вычисление ошибки сигмы приводится в разделе «а» § 1.10, а ошибка и оценка M — в § 1.03.

в. Способ распределения дат.

Излагаемую схему расчета удобно применять, если все варианты выборки — малозначные целые числа и среди них много повторяющихся. Формула для вычисления сигмы по способу распределения дат;

$$\sigma = \pm \sqrt{\frac{\sum fx^2 - (\sum fx)^2/N}{N-1}}, \quad (1.38)$$

где σ — среднее квадратическое отклонение; $\sum fx^2$ — сумма произведений частот на квадраты вариантов; $\sum fx$ — сумма произведений частот на варианты; N — объем выборки.

Порядок действий (на примере данных из табл. 1.01).

1. Выписываем поочередно в столбец 1 табл. 1.14 все варианты и отмечаем в столбце 2, сколько раз каждая из них встретилась в выборке.

2. Умножаем варианты на их частоты (столбец 3).

3. Числа столбца 3 множим на числа столбца 1, результаты записываем в столбце 4.

4. Суммируем числа каждого столбца, кроме первого, и получаем:

$$N = 214; \quad \Sigma fx = 16069; \quad \Sigma fx^2 = 1214241.$$

5. Подставляем полученные суммы в формулу (1.38):

$$\sigma = \sqrt{\frac{1214241 - 16069^2/214}{214 - 1}} = 5,9\bar{9}.$$

Одновременно может быть вычислена средняя арифметическая по формуле (1.13):

$$M = 16069/214 = 75,09.$$

Вычисление ошибки среднего квадратического отклонения — см. раздел «а» § 1.10, а ошибки M — см. § 1.03.

Одним из достоинств способа распределения дат является также то, что в этом случае отпадает необходимость введения так называемой поправки Шеппарда к величине сигмы.

г. Способ произведений.

Этот способ вычисления сигмы отличается от предыдущего тем, что здесь составляется взвешенный вариационный ряд с равными классовыми интервалами, а по способу «в» интервалы могут быть неравными. Напомним, что наиболее простым по технике вычислительной работы является способ «б», который и рекомендуется при симметричных распределениях в большинстве случаев. Для вычисления сигмы по способу произведений используется формула (1.38).

Порядок работы (табл. 1.15).

1. Перемножая частоты из столбца 2 на соответствующие середины классов из столбца 1, получаем числа столбца 3 таблицы: 120, 1950 и т. д.

2. Числа столбца 3 последовательно множим на варианты:

$$120 \cdot 60 = 7200; \quad 1950 \cdot 65 = 126750 \text{ и т. д.}$$

3. Суммируя числа каждого столбца, кроме первого, получаем:

$$N = 214; \quad \Sigma fx = 16060; \quad \Sigma fx^2 = 1213100.$$

4. По формуле (1.38) вычисляем:

$$\sigma = \sqrt{\frac{1213100 - 16060^2/214}{214 - 1}} = 6,07.$$

Средняя арифметическая в данном случае может быть получена по формуле (1.13):

$$M = 16060/214 = 75,05.$$

Таблица 1.14. К вычислению среднего квадратического отклонения по способу распределения вариант

Варианта, x	Частота, f	fx	fx^2
1	2	3	4
92	2	184	16 928
91	1	91	8 281
88	1	88	7 744
86	1	86	7 396
84	6	504	42 336
83	1	83	6 889
82	11	902	73 964
81	7	567	45 927
80	23	1 840	147 200
79	12	948	74 892
78	21	1 638	127 764
77	14	1 078	83 006
76	14	1 064	80 864
75	14	1 050	78 750
74	13	962	71 188
73	8	584	42 632
72	11	792	57 024
71	6	426	30 246
70	5	350	24 500
69	5	345	23 805
68	7	476	32 368
67	6	402	26 934
66	5	330	21 780
65	8	520	33 800
64	7	448	28 672
63	3	189	11 907
62	1	62	3 844
60	1	60	3 600
214	16 069	1 214 241	

Таблица 1.15. К вычислению сигмы по способу произведений

Варианта (середина класса), x	Частота, f	fx	fx^2
60	2	120	7 200
65	30	1 950	126 750
70	34	2 380	166 600
75	62	4 650	348 750
80	74	5 920	473 600
85	8	680	57 800
90	4	360	32 400
214	16 060	1 213 100	

Таблица 1.16. К вычислению сигмы по способу случайных серий

Серия	x_N		$x_N - x_1$
1	77	63	14
2	84	65	19
3	84	64	20
4	91	69	22
5	86	78	8
6	84	75	9
7	80	75	5
8	81	70	11
9	82	69	13
10	80	64	16
11	77	60	17
12	92	65	27
			181

д. Прямой способ.

Для машинного счета удобна при вычислении сигмы формула (1.24), написанная в применении к взвешенному ряду:

$$\sigma = + \sqrt{\frac{\sum (x - M)^2 f}{N - 1}}, \quad (1.38a)$$

где σ — среднее квадратическое отклонение; $\sum (x - M)^2 f$ — сумма произведений квадратов отклонений вариант от средней на соответствующие частоты; N — объем выборки. Напомним, что формула $\sum (x - M) f = 0$ в пределах точности расчетов служит для проверки вычислений средней арифметической и отклонений от нее.

Пример обработки ряда по этому способу дан в § 1.34. Определение ошибки сигмы рассматривается в разделе «а» § 1.10, а расчет ошибки и оценка M приводятся в § 1.03.

§ 1.11. Приближенное вычисление среднего квадратического отклонения по способу случайных серий, размаху варьирования и коэффициенту вариации

При планировании эксперимента может возникнуть необходимость определить, хотя бы приблизительно, объем выборки, достаточный для получения достоверных результатов или выводов по тому или иному вопросу. Для этого бывает достаточно вычислить приближенное значение среднего квадратического отклонения по способу случайных серий, размаху варьирования данных и по величине коэффициента вариации, что значительно облегчает расчеты, и в то же время позволяет получить значение сигмы, достаточно точное для названной цели. Вычисление сигмы по способу случайных серий на примере данных табл. 1.01 проводится следующим образом.

1. Неранжированная выборка разбивается на случайные серии, по 9 вариант в каждой так, чтобы в эти серии попали максимальная и минимальная варианты (если выборка берется неполностью). Если выборка берется полностью и число ее вариант не кратно 9, то некоторые серии могут налегать краями друг на друга так, что их крайние варианты будут общими для граничащих серий. Левая половина табл. 1.01 образует 12 строк, которые примем за случайные серии по 9 вариант в каждой: сверху вниз первая строка — первая серия, вторая строка — вторая серия.

2. Подсчитаем размах вариации в сериях, для чего из максимальной варианты каждой серии вычитаем минимальную варианту (столбец 4 табл. 1.16).

3. Средний размах получается от деления суммы размахов (181) всех серий на их число: $n = 12$.

4. Полученное число делим на 3 ($3 = \sqrt{9}$). Если бы вместо девяток разбивали совокупность на четверки, то делили бы не на 3, а на 2 ($2 = \sqrt{4}$).

Указанные действия отображены в формуле:

$$\sigma = \frac{\sum (x_N - x_1)}{3n} \quad (1.39)$$

где σ — среднее квадратическое отклонение; x_N — максимальная варианта каждой серии; x_1 — минимальная варианта серии; n — число серий.

Для данных табл. 1.01 и 1.16:

$$\sigma = \frac{181}{3 \cdot 12} = 5,03.$$

Точное значение сигмы для вегетационного периода ячменя равно 5,989 (§ 1.10, в).

При нормальном распределении вариант весь размах варьирования ($x_N - x_1$) занимает около 6σ . На этом основан приближенный

метод вычисления сигмы по размаху варьирования по формуле:

$$\sigma = \frac{x_N - x_1}{6}, \quad (1.40)$$

или по формуле:

$$\sigma = \frac{x_N - x_1}{h}, \quad (1.41)$$

где σ — среднее квадратическое отклонение; x_N — максимальная варианта выборки; x_1 — минимальная варианта; h — коэффициент, выбираемый по табл. 4П.

Для ряда из табл. 1.02 по табл. 4П $h = 3,5$; по формуле (1.41): $\sigma = (82 - 43)/3,5 = 11,14$. Точное значение сигмы этого ряда: 11,976 (см. § 1.09, б).

В тех случаях, когда при планировании опыта известно предполагаемое значение средней арифметической данной совокупности и примерная, хотя бы определенная глазомерно, величина коэффициента вариации (см. о нем в § 1.13), можно определить приближенно среднее квадратическое отклонение для установления минимального в данном эксперименте объема выборки по преобразованной формуле (1.43):

$$\sigma = \frac{Mv}{100}$$

где σ — среднее квадратическое отклонение; M — средняя арифметическая; v — приблизительное значение коэффициента вариации.

Например, предполагаемая средняя длина листьев традесканции: $M = 40$ мм и по глазомерной оценке $v = 10\%$. Отсюда:

$$\sigma = \frac{40 \cdot 10}{100} = 4 \text{ мм.}$$

Найденное значение сигмы можно также использовать для определения требуемой численности выборки при планировании эксперимента по формуле (1.102).

§ 1.12. Объединение выборок по их дисперсиям

В § 1.05 было показано, как объединить выборки по их средним арифметическим. Часто при этом требуется также найти и среднюю дисперсию объединенной выборки, которую можно вычислить по формуле

$$\sigma^2 = \frac{\sum (N_i - 1) \sigma_i^2}{\sum N_i - k}, \quad (1.42)$$

где σ^2 — средняя дисперсия; σ_i^2 — дисперсии частных выборок; N_i — объемы частных выборок; k — число частных выборок.

Вычислим среднюю дисперсию для четырех групп измерений диаметра цветков гелениума осеннего: $\sigma_1^2 = 2,00$; $N_1 = 7$; $\sigma_2^2 = 3,35$; $N_2 = 5$; $\sigma_3^2 = 2,95$; $N_3 = 8$; $\sigma_4^2 = 4,37$; $N_4 = 9$.

По формуле (1.42)

$$\sigma^2 = \frac{(7-1)2,00 + (5-1)3,35 + (8-1)2,95 + (9-1)4,37}{7+5+8+9-4} = 3,24,$$

откуда: $\sigma = \sqrt{3,24} = 1,80$.

§ 1.13. Коэффициент вариации

Среднее квадратическое отклонение характеризует степень отклонения вариант данной совокупности от среднего арифметического в абсолютных числах. Однако для сравнения вариабельности двух или более совокупностей величины сигмы не могут быть использованы, за исключением редкого случая, когда средние арифметические одинаковы или очень близки друг к другу. Во всех остальных случаях для сравнения совокупностей по их вариабельности необходимо вычислять коэффициент вариации.

Коэффициент вариации показывает, какой процент составляет сигма от средней арифметической, и позволяет сравнивать между собой по степени варьирования любые совокупности. Коэффициент вариации имеет важное значение в биологии и агрономии, например для установления степени выровненности популяции или сорта по тому или иному признаку.

Для чисел с одинаковым знаком коэффициент вариации вычисляют по формуле

$$v = \frac{\sigma}{M} \cdot 100\%, \quad (1.43)$$

где v — коэффициент вариации; M — средняя арифметическая; σ — среднее квадратическое отклонение.

Для данных о длине вегетационного периода сортов ячменя (табл. 1.03 и § 1.10):

$$v = \frac{6,06 \cdot 100}{75,047} = 8,1\%.$$

Следовательно, изученные сорта ячменя сравнительно мало варьируют по продолжительности вегетационного периода.

Для данных по высоте растений сои (табл. 1.02 и § 1.09):

$$v = \frac{11,976 \cdot 100}{64,538} = 18,6\%.$$

Растения сои имеют среднее варьирование данных по высоте или, точнее, относятся ко второму баллу вариабельности по шести балльной шкале (см. § 1.37).

Иногда при вычислении ошибки средней арифметической, например по формуле (1.11), значение сигмы не фиксируют. Величину коэффициента вариации в таких случаях можно найти по формуле:

$$v = \frac{100m_M \sqrt{N}}{M}, \quad (1.44)$$

где m_M — ошибка средней арифметической; N — объем совокупности; M — средняя арифметическая.

Если требуется найти коэффициент вариации, минуя вычисление сигмы, то можно применить формулу:

$$v = 100 \sqrt{\frac{\frac{\sum x^2}{N} - M^2}{N - 1}}. \quad (1.45)$$

Например, по данным табл. 1.11, $\sigma = 11,98$; $M = 64,54$; $\sum x^2 = 55869$; $N = 13$:

$$v = 100 \sqrt{\frac{55869/64,54^2 - 13}{13 - 1}} = 19\%,$$

что практически совпадает с результатом по формуле (1.43).

Ошибка коэффициента вариации определяется по формуле:

$$m_v = v \sqrt{\frac{0,5 + 0,0001v^2}{N}}, \quad (1.46)$$

где m_v — ошибка коэффициента вариации; v — коэффициент вариации; N — объем выборки.

Чаще для вычисления ошибки коэффициента вариации пользуются менее точной формулой:

$$m_v = \frac{v}{\sqrt{2N}}, \quad (1.47)$$

где обозначения те же.

Для рассматриваемого примера о продолжительности вегетационного периода ячменя ошибка его коэффициента вариации по формуле (1.46):

$$m_v = 8,1 \sqrt{\frac{0,5 + 0,0001 \cdot 8,1^2}{214}} = \pm 0,394\%,$$

а по формуле (1.47):

$$m_v = \frac{8,1}{\sqrt{2 \cdot 214}} = \pm 0,391\%.$$

Оба результата практически совпадают. Следовательно, при отсутствии особо высоких требований к точности определения ошибки коэффициента вариации можно пользоваться формулой (1.47). О вычислении коэффициента вариации совокупности с альтернативным разделением вариантов см. § 1.26. Если прибавить ко всем вариантам ряда одно и тоже число n , или вычесть его из них, то величина коэффициента вариации изменится по формуле:

$$v = \frac{100\sigma}{M \pm n} \quad (1.48)$$

Если данные представляют собой смесь отрицательных и положительных чисел, как, например, это бывает при обработке резуль-

татов измерения температуры, то коэффициент вариации следует вычислять по формуле:

$$v = \frac{100\sigma}{|x_1| + M}, \quad (1.49)$$

где v — коэффициент вариации; σ — среднее квадратическое отклонение; $|x_1|$ — наименьшая отрицательная варианта ряда без знака; M — средняя арифметическая.

Смысл этой формулы обусловлен тем, что при вычислении коэффициента вариации средняя арифметическая, так же как и сигма, должна быть представлена в виде отрезка на числовой оси.

Алгоритм вычисления коэффициента вариации для относительных или рациональных чисел покажем на следующем примере. Для ряда из 6 чисел: $-3, -2, -1, 0, +1, +2$ средняя арифметическая равна $-0,5$, а сигма $1,87$. Следовательно, по формуле (1.43) было бы:

$$v = \frac{1,87 \cdot 100}{-0,5} = -374\%,$$

чего не может быть для такого небольшого варьирования данных. Если же M измерить как отрезок от точки -3 до точки $-0,5$, то она равна: $|-3| + (-0,5) = 2,5$, а по формуле (1.49) коэффициент вариации равен:

$$v = \frac{100 \cdot 1,87}{|-3| + (-0,5)} = 75\%,$$

что гораздо ближе соответствует реальности.

В § 1.37 рассмотрена аппроксимация логнормальным распределением данных по варьированию 165 различных признаков растений и предложена шкала в процентах: 1) небольшое варьирование (от 0 до 4); 2, 3) нормальное (от 5 до 44); 4) большое (от 45 до 64); 5) очень большое (от 65 до 84); 6) сверхбольшое (от 85 до 104); 7) аномальное (от 105 и больше). В пределах интервала нормы варьирование может различаться: 2) от 5 до 24 — «нижняя» норма; 3) от 25 до 44 — «верхняя» норма. В последнем примере варьирование относится к пятому классу и может считаться очень большим.

§ 1.14. Показатель точности опыта

Показатель точности опыта выражает величину ошибки средней арифметической в процентах от самой средней арифметической и, таким образом, служит показателем точности определения последней. Чем больше показатель P , т. е. процент ошибки, тем меньше точность опыта. Показатель точности опыта вычисляется по формуле:

$$P = \frac{m_M}{M} \cdot 100\%, \quad (1.50)$$

или по формуле:

$$P = \frac{v}{\sqrt{N}}, \quad (1.51)$$

где P — показатель точности опыта; M — средняя арифметическая; m_M — ошибка средней арифметической; v — коэффициент вариации; N — объем выборки.

Для данных по вегетационному периоду ячменя $M = 75,05$ и ошибка $m_M = \pm 0,415$.

Отсюда по формуле (1.50) показатель точности опыта:

$$P = \frac{0,415 \cdot 100}{75,05} = 0,55\%.$$

Ошибка показателя точности опыта вычисляется по формуле:

$$m_P = \pm P \sqrt{\frac{1}{2N} + \left(\frac{P}{100}\right)^2} \quad (1.52)$$

где m_P — ошибка показателя точности опыта; P — показатель точности опыта; N — объем выборки.

В данном случае:

$$m_P = \pm 0,55 \sqrt{\frac{1}{2 \cdot 214} + \left(\frac{0,55}{100}\right)^2} = \pm 0,0268.$$

В результате $P = 0,55 \pm 0,03\%$.

Точность опыта считается удовлетворительной, если величина показателя не превышает 5%. При значениях больше 5%, рекомендуется увеличить число наблюдений или повторностей. Иногда можно снизить величину P , если повысить точность измерений объектов опыта. О вычислении показателя точности опыта при дисперсионном анализе см. § 4.16.

§ 1.15. Медиана

Варианта, которая разделяет ряд на две равные по числу вариант части, называется медианой. Порядковый номер медианной варианты от начала ранжированного ряда можно определить по выражению: $0,5(N + 1)$. Используется медиана как средний показатель описательного типа или в качестве средней положения и для непараметрического теста медианы (см. § 4.08). Вычисляется медиана взвешенного ряда по формуле:

$$M_e = x_e + c \left(\frac{0,5N - L}{f_e} \right), \quad (1.53)$$

где M_e — медиана; x_e — начало класса, в котором находится медиана; c — величина классового интервала; N — объем выборки; L — сумма частот классов (начиная с меньшего), предшествующих классу, в котором находится медиана; f_e — частота медианного класса.

Рассчитаем медиану вегетационного периода совокупности сортов ячменя (по данным табл. 1.03). Порядковый номер медианной варианты: $0,5(214 + 1) = 107,5$. Класс, в котором находится медиана, определяется путем накопления частот. Так, в табл. 1.03 сумма частот сверху вниз равна: $2 + 30 + 34 + 62 = 128$. Поскольку между 107 и 128 вариантами должна находиться медиана, накопле-

ние частот прекращаем и за начало медианного класса принимаем 72,5, т. е. меньшую границу класса: 72,5—77,4, среднее значение которого 75, а частота 62. Отсюда по формуле (1.53):

$$M_e = 72,5 + 5 \left(\frac{0,5 \cdot 214 - 66}{62} \right) = 75,8 \text{ дня.}$$

Это означает, что в данной совокупности сортов ячменя половина из них (107 сортов) созревает не позже, чем на 76-й день от начала вегетационного периода, а вторая половина сортов имеет периоды созревания больше 76 дней.

Ошибка медианы вычисляется сравнительно редко. Для достаточно больших рядов определить ее можно по формуле:

$$m_e = 1,25331 \frac{\sigma}{\sqrt{N}}, \quad (1.54)$$

где m_e — ошибка медианы; σ — среднее квадратическое отклонение; N — объем выборки; $\frac{\sigma}{\sqrt{N}}$ — ошибка средней арифметической (см. § 1.03).

Значение константы в формуле (1.54) зависит от объема ряда; для некоторых четных N от 2 до 20 оно равно: 2—1,0; 4—1,092; 6—1,135; 8—1,160; 10—1,177; 12—1,189; 20—1,214; ∞ — 1,253. В рассмотренном примере:

$$m_e = 1,25331 \cdot \frac{5,189}{\sqrt{214}} = 0,5 \text{ дня.}$$

Таким образом, медиана равна $75,8 \pm 0,5$ и по своей величине близка к средней арифметической ($M = 75,05$), что может служить одним из доводов в пользу заключения о близости данного распределения к симметричному.

§ 1.16. Вычисление 11 параметров невзвешенного ряда

В этом параграфе приводится схема совокупного вычисления всех основных статистических параметров невзвешенного ряда наиболее экономичным способом в расчете на то, что исследователь располагает лишь простейшими вычислительными средствами: счетами, арифмометром и таблицей квадратных корней.

Основными параметрами выборки считаются средняя арифметическая и среднее квадратическое отклонение (сигма), так как почти все прочие показатели невзвешенного ряда вычисляются на основании их числовых значений.

Первое место по трудоемкости вычислений среди обычных статистик занимает сигма. Поэтому разработано несколько способов, упрощающих ее расчет. Наименее утомительным является способ условной средней, который значительно упрощает вычисление сигмы и обеспечивает любую требуемую точность.

Исходные данные для приводимого примера расчета являются датами зацветания ольхи кустарниковой в Ленинграде за 14 лет

Таблица 1.17. К вычислению параметров невзвешенного ряда способом условной средней

Даты зацветания, в днях от 1 марта (см. табл. 24П)	$a = x - A$		Даты зацветания, в днях от 1 марта (см. табл. 24П)	$a = x - A$	
65	-10	100	73	-2	4
67	-8	64	78	+3	9
69	-6	36	81	+6	36
70	-5	25	85	+10	100
71	-4	14	86	+11	121
72	-3	9	87	+12	144
72	-3	9			
73	-2	4			
			1049	-1	677

(табл. 1.17). Вычисления проводятся в следующей последовательности.

1. Суммируем варианты ряда и делим их на число наблюдений, получаем среднюю арифметическую:

$$M = \frac{\sum x}{N} = \frac{1049}{14} = 74,928.$$

2. Ближайшее к M целое число $A = 75$ используем для получения условных отклонений по формуле $a = x - A$, приведенных в столбце 2. Суммируем отклонения с учетом их знаков $\sum a = -1$. Делаем проверку вычислений по формуле: $\sum a = \sum x - NA$; $-1 = 1049 - 14 \cdot 75$; $-1 = -1$; равенство обеих частей уравнения соблюдается; следовательно, вычисления сделаны верно.

3. Возводим в квадрат условные отклонения (столбец 3) и суммируем эти квадраты: $\sum a^2 = 677$.

4. Определяем дисперсию по формуле:

$$\sigma^2 = \frac{\sum a^2 - (\sum a)^2 \cdot N}{N - 1} = \frac{677 - (-1)^2 \cdot 14}{14 - 1} = 52,07,$$

откуда $\sigma = \sqrt{52,07} = 7,2$.

5. Прочие параметры данного статистического ряда определим следующим образом.

Ошибка средней арифметической:

$$m_M = \frac{\sigma}{\sqrt{N}} = \frac{7,2}{\sqrt{14}} = \pm 1,9.$$

Критерий достоверности средней арифметической:

$$t_M = \frac{M}{m_M} = \frac{74,9}{1,9} = 39,4.$$

Ошибка сигмы:

$$m_\sigma = \frac{\sigma}{\sqrt{2N}} = \frac{7,2}{\sqrt{2 \cdot 14}} = \pm 1,4.$$

Коэффициент вариации:

$$v = \frac{100\sigma}{M} = \frac{100 \cdot 7,2}{74,9} = 9,6\%.$$

Ошибка коэффициента вариации:

$$m_v = v \sqrt{\frac{0,5 + 0,0001v^2}{N}} = 9,6 \sqrt{\frac{0,5 + 0,0001 \cdot 9,6^2}{14}} = \pm 1,8\%.$$

Показатель точности опыта:

$$P = \frac{m_M}{M} \cdot 100 = \frac{1,9}{74,9} \cdot 100 = 2,5\%.$$

Ошибка показателя точности опыта:

$$m_P = P \sqrt{\frac{1}{2N} + \left(\frac{P}{100}\right)^2} = 2,5 \sqrt{\frac{1}{2 \cdot 14} + \left(\frac{2,6}{100}\right)^2} = 0,5\%.$$

Медиана определяется следующим образом. Находим номер медианной варианты, который равен $0,5 (N + 1) = 0,5 (14 + 1) = 7,5$. При четном числе вариантов за медиану принимается середина промежутка между двумя центральными вариантами. При нечетном числе вариантов за медиану принимается центральная варианта. В нашем случае, при четном числе вариантов, медиана равна полусумме вариантов 7 и 8: $M_e = (72 + 73)/2 = 72,5$.

Таким образом, получено 11 параметров, которые дают довольно полное представление о данном статистическом невзвешенном ряде. Истолковать эти числа можно примерно так. В среднем за 14 лет ольха кустарниковая в Ленинграде зацветает через 74,9 дня от 1 марта, т. е. (см. табл. 24П) округленно 14 мая, с ошибкой 1,9 дня, и с доверительным интервалом на 0,95 уровне (см. § 1.23): $13,9 \pm 2 \cdot 1,9$; т. е. самый ранний срок зацветания 10 мая, и самый поздний — 18 мая. Варьирование годовых дат зацветания сравнительно невелико, так как коэффициент вариации равен: $v = 9,6 \pm 1,8\%$. Полученные параметры заслуживают доверия ввиду большой достоверности средней арифметической: $t_M = 38,9 \gg 3$ и значения показателя точности опыта, меньшего 5%: $P = (2,5\% \pm 0,5\%) < 5\%$.

§ 1.17. Мода

Модой называется точка на оси абсцисс, соответствующая максимальной частоте теоретической кривой распределения вариантов. Мода вычисляется лишь тогда, когда вариационный ряд разбит на классы, содержащие разное число частот. В случаях, не требующих особой точности, за моду можно принять приближенно наиболее часто встречающуюся варианту ряда. Например, в ряду из табл. 1.03 за моду можно принять $M_o = 75$, т. е. середину класса 72,5—77,4, который имеет наибольшую частоту — 62. Положение моды относительно средней арифметической определяет степень скошенности, или асимметрии, кривой распределения данного ряда. Мода в симметричном нормальном распределении совпадает со средней арифме-

тической. Вычисляется мода по формуле:

$$M_o = x_o + c \left(\frac{f_2 - f_1}{2f_2 - f_1 - f_3} \right), \quad (1.55)$$

где M_o — мода; x_o — начало модального класса, т. е. меньшая граница того класса, который имеет наибольшую частоту; c — величина классового интервала; f_1 — частота класса, предшествующего модальному; f_2 — частота модального класса; f_3 — частота класса, следующего за модальным.

Определим моду вегетационного периода совокупности сортов ячменя (табл. 1.03) по формуле (1.55):

$$M_o = 77,5 + 5 \left(\frac{74 - 62}{2 \cdot 74 - 62 - 8} \right) = 78,3.$$

Это означает, что наиболее часто встречаются сорта ячменя, зерно у которых созревает на 78-й день от начала их вегетационного периода.

Ошибка моды обычно бывает больше ошибки средней арифметической той же совокупности. Ввиду сложности и малой целесообразности ее вычисления она здесь не приводится. Наличие нескольких мод (пиков) в вариационном ряду большей частью является указанием на неоднородность совокупности. В невзвешенном вариационном ряду мода отсутствует.

§ 1.18. Условные, центральные и факториальные моменты распределения

Моментами распределения называют средние степени отклонений вариант от средней арифметической или от произвольного числа, в частности от нуля. В первом случае моменты называются центральными, во втором — условными, в третьем — начальными. Порядок момента равен степени, в которую возводятся отклонения. Основным моментом называется частное от деления центрального момента соответствующего порядка на сигму, возведенную в степень, равную тому же порядку момента. Моменты, особенно центральные, играют большую роль при вычислении различных показателей вариационного ряда. Средняя арифметическая — это начальный момент первого порядка, а дисперсия — центральный момент второго порядка. Условные и факториальные моменты большей частью вычисляют в качестве вспомогательных, для получения центральных моментов. Приведенные ниже вычисления моментов имеют главным образом методическое значение, они выполнены для иллюстрации самого понятия моментов. В практических расчетах рекомендуется для вычисления моментов применять способ кодированных вариантов (§ 1.22), а также способы сумм, рассмотренные в конце данного параграфа.

В табл. 1.18 приведена сводка таких формул вычисления моментов, которые соответствуют приведенным выше определениям последних. Обозначения моментов первых четырех порядков в табл. 1.18 следующие: $\mu_1, \mu_2, \mu_3, \mu_4$ — центральные; m_1, m_2, m_3, m_4 — услов-

Таблица 1.18. Формулы для вычисления моментов распределения

Порядок моментов распределения	Центральные	Условные	Начальные	Основные
Первый	$\mu_1 = \frac{\Sigma (x - M) f}{N} = 0$	$m_1 = \frac{\Sigma (x - A) f}{N}$	$b_1 = \frac{\Sigma (x - 0) f}{N} = M$	$r_1 = \frac{\mu_1}{\sigma} = 0$
Второй	$\mu_2 = \frac{\Sigma (x - M)^2 f}{N} = \sigma^2$	$m_2 = \frac{\Sigma (x - A)^2 f}{N}$	$b_2 = \frac{\Sigma (x - 0)^2 f}{N}$	$r_2 = \frac{\mu_2}{\sigma^2} = 1$
Третий	$\mu_3 = \frac{\Sigma (x - M)^3 f}{N}$	$m_3 = \frac{\Sigma (x - A)^3 f}{N}$	$b_3 = \frac{\Sigma (x - 0)^3 f}{N}$	$r_3 = \frac{\mu_3}{\sigma^3} = A_s$
Четвертый	$\mu_4 = \frac{\Sigma (x - M)^4 f}{N}$	$m_4 = \frac{\Sigma (x - A)^4 f}{N}$	$b_4 = \frac{\Sigma (x - 0)^4 f}{N}$	$r_4 = \frac{\mu_4}{\sigma^4}$

Таблица 1.19. Вычисление условных моментов

Варианта, x	Частота, f	$\frac{x - A}{(A = 75)}$	$(x - A) f$	$(x - A)^2$	$(x - A)^2 f$	$(x - A)^3$	$(x - A)^3 f$	$(x - A)^4$	$(x - A)^4 f$
60	2	-15	-30	225	450	-3375	-6 750	50 625	101 250
65	30	-10	-300	100	3000	-1000	-30 000	10 000	300 000
70	34	-5	-170	25	850	-125	-4 250	625	21 250
75	62	0	0	0	0	0	0	0	0
80	74	+5	370	25	1850	125	9 250	625	46 250
85	8	+10	80	100	800	1000	8 000	10 000	80 000
90	4	+15	60	225	900	3375	13 500	50 625	202 500
c = 5	214	0	+10	—	7850	—	-10 250	—	751 250

ные; b_1, b_2, b_3, b_4 — начальные; r_1, r_2, r_3, r_4 — основные¹. Остальные обозначения: Σ — знак суммирования; x — варианты; f — частоты вариантов; N — объем выборки; M — средняя арифметическая; A — начальная точка отчета, или условная средняя; σ — среднее квадратическое отклонение (сигма); A_s — показатель асимметрии (см. § 1.19).

Центральные моменты можно вычислить прямым способом по формулам, указанным в табл. 1.18, т. е. через отклонения вариантов от средней арифметической, но предпочитают их получать через условные моменты, для чего рекомендуются следующие формулы:

$$\mu_2 = m_2 - m_1^2; \quad (1.56)$$

$$\mu_3 = m_3 + m_1 (2m_1^2 - 3m_2); \quad (1.57)$$

$$\mu_4 = m_4 + m_1 [3m_1 (2m_2 - m_1^2) - 4m_3], \quad (1.58)$$

где обозначения те же, что и к табл. 1.18. Проверка вычислений центральных моментов производится по формулам:

$$\mu_3 = m_3 - m_1 (3\mu_2 + m_1^2); \quad (1.59)$$

$$\mu_4 = m_4 - m_1 [4\mu_3 + m_1 (6\mu_2 + m_1^2)]. \quad (1.60)$$

На примере данных из табл. 1.03 вычисление условных моментов показано в табл. 1.19, подставляя суммы столбцов из которой в формулы условных моментов из табл. 1.18, получаем:

$$m_1 = 10/214 = 0,0467; \quad m_2 = 7850/214 = 36,682;$$

$$m_3 = -10250/214 = -47,897; \quad m_4 = 751250/214 = 3510,5.$$

Подставив значения условных моментов в формулы (1.56) — (1.58), определим величины центральных моментов:

$$\mu_2 = 36,682 - 0,0467^2 = 36,680;$$

$$\mu_3 = -47,897 + 0,0467 (2 \cdot 0,0467^2 - 3 \cdot 36,682) = -53,036;$$

$$\mu_4 = 3510,5 + 0,0467 [3 \cdot 0,0467 (2 \cdot 36,682 - 0,0467^2) + 4 \cdot 47,897] = 3519,9.$$

Проверку вычисления центральных моментов делаем по формулам (1.59), (1.60):

$$\mu_3 = -47,8972 - 0,0467 (3 \cdot 36,680 + 0,0467^2) = -53,036;$$

$$\mu_4 = 3510,5 - 0,0467 [-4 \cdot 53,039 + 0,0467 (6 \cdot 36,680 + 0,0467^2)] = 3519,9.$$

Если после проверки не будет достигнуто совпадение величин центральных моментов в пределах точности вычислений, то следует искать ошибку в расчете.

¹ В главе 2 через r будет обозначаться также коэффициент корреляции. Если r означает моменты, то при нем стоит подстрочный индекс в виде цифры 3 или 4, а при r , означающем коэффициент корреляции, подстрочные индексы иные, или они отсутствуют.

По данным столбцов 1 и 2 табл. 1.19 вычислим также начальные моменты по формулам из табл. 1.18:

$$b_1 = 16060/214 = 75,05; \quad b_2 = 1213100/214 = 5668,69;$$

$$b_3 = 92206000/214 = 430869,16;$$

$$b_4 = 7050582500/214 = 32946647,19.$$

В биометрии из начальных моментов, которые есть разновидность условных, применяется обычно лишь момент первого порядка, который равен средней арифметической; начальный момент второго порядка иногда используется при вычислении среднего квадратичного отклонения.

Из основных моментов вычисляются лишь r_3 , который является коэффициентом асимметрии (см. § 1.19), и r_4 , который служит для вычисления эксцесса (см. § 1.20). Оба этих основных момента часто применяют для расчета теоретических частот некоторых типов кривых распределения.

Хотя моменты и не имеют самостоятельного интереса для биометрического анализа, они необходимы для вычисления почти всех параметров взвешенного вариационного ряда, поэтому должны быть вычислены достаточно точно. Их рекомендуется вычислять наиболее рациональным способом кодированных вариантов, приведенным в § 1.22.

Когда варианты в совокупности распределяются симметрично, или хотя бы по двускатной кривой, начальные и центральные моменты, вычисленные, например, по формулам из табл. 1.18, а также (1.56)—(1.58), могут быть уточнены поправками Шеппарда. Посредством моментов вычисляется большинство параметров статистического ряда, поэтому внесение поправок Шеппарда к моментам отразится далее почти на всех данных, получаемых в ходе статистического исследования. Эти поправки вводятся для компенсации неточности определения моментов именно по взвешенному ряду, когда приходится варианты разносить по классам и тем самым как бы округлять их до среднего значения того класса, в который они попали. Поэтому поправки Шеппарда не нужны, когда моменты вычисляются непосредственно по исходным данным, без группировки их во взвешенный ряд с равными интервалами. Например, не требуется поправка к значению среднего квадратического отклонения, вычисленного в § 1.10 по способу распределения дат с естественной группировкой последних.

Поправки Шеппарда к начальным и центральным моментам вычисляются по формулам:

$$\tilde{b}_2 = b_2 - \frac{c^2}{12}; \quad (1.61)$$

$$\tilde{b}_3 = b_3 - \frac{b_1 c^2}{4}; \quad (1.62)$$

$$\tilde{b}_4 = b_4 - \frac{c^2}{2} \left(b_2 + \frac{7c^2}{120} \right), \quad (1.63)$$

где $\tilde{b}_2 - \tilde{b}_4$ — исправленные начальные моменты; $b_1 - b_4$ — исходные, неисправленные начальные моменты (формулы их вычисления

см. табл. 1.18); c — классовой интервал.

$$\tilde{\mu}_2 = \mu_2 - \frac{c^2}{12}; \quad (1.64)$$

$$\tilde{\mu}_4 = \mu_4 - \frac{c^2}{2} \left(\mu_2 - \frac{7c^2}{120} \right), \quad (1.65)$$

где $\tilde{\mu}_2, \tilde{\mu}_4$ — исправленные центральные моменты; μ_2, μ_4 — неисправленные центральные моменты, вычисленные по формулам табл. 1.18 и (1.56)—(1.58).

Третий центральный момент μ_3 поправок не требует.

Введем поправки в вычисленные моменты:

$$b_1 = 75,05; \quad b_2 = 5668,69; \quad b_3 = 430869,16; \quad b_4 = 32946647,19; \\ \mu_2 = 36,68; \quad \mu_4 = 3519,95; \quad c = 5.$$

Исправленные значения моментов по формулам (1.61)—(1.65) следующие:

$$\tilde{b}_2 = 5668,69 - \frac{5^2}{12} = 5666,61,$$

$$\tilde{b}_3 = 430869,16 - \frac{75,05 \cdot 5^2}{4} = 430400,10,$$

$$\tilde{b}_4 = 32946647,19 - \frac{5^2}{2} \left(5668,69 - \frac{7 \cdot 5^2}{120} \right) = 32875806,79,$$

$$\tilde{\mu}_2 = 36,68 - \frac{5^2}{12} = 34,60,$$

$$\tilde{\mu}_4 = 3519,95 - \frac{5^2}{2} \left(36,68 - \frac{7 \cdot 5^2}{120} \right) = 3079,68.$$

Как видим, в некоторых случаях поправки Шеппарда могут заметно влиять на величину моментов, однако следует иметь в виду, что применение поправок к несоответствующим распределениям может внести также и заметную ошибку, поэтому единого мнения о необходимости введения поправок Шеппарда пока не имеется.

Факториальные моменты ввиду простоты их получения в прикладных вычислениях применяются для расчетов условных, далее центральных моментов, среднего квадратического отклонения и средней арифметической.

В этом параграфе приводятся пять способов вычисления условных и центральных моментов посредством факториальных. Некоторые из этих способов дают различные величины факториальных и условных моментов, однако центральные моменты всегда при этом совпадают. Для сравнимости результатов и облегчения понимания сути приводимых алгоритмов все пять способов расчета выполнены на примере одного и того же взвешенного ряда распределения величин диаметра соцветий (в мм) у нивяника.

а. Способ произведений.

Вычисление факториальных моментов по этому способу приводится здесь для иллюстрации принципа их построения, а также ввиду того, что формулы способа произведений в общем случае более

Таблица 1.20. Вычисление факториальных моментов по способу произведений ряда распределения диаметра соцветий нивяника (x)

	f		fa_1	fa_2	fa_3	fa_4
37	2	—3	—6	24	—120	720
51	29	—2	—58	174	—696	3480
65	119	—1	—119	238	—714	2856
79	171	0	0	0	0	0
93	121	1	121	0	0	0
107	42	2	84	84	0	0
121	16	3	48	96	96	0
$c = 14 \quad N = 500$			$s_1 = 70 \quad s_2 = 616 \quad s_3 = -1434 \quad s_4 = 7056$			

удобны для программирования на ЭВМ. При ручном счете факториальные моменты рекомендуется вычислять по одному из приведенных далее в этом параграфе способов сумм.

Последовательность расчетов факториальных моментов по способу произведений следующая.

1. По данным распределения соцветий нивяника по диаметру записываем середины классов x и частоты f в табл. 1.20.

2. Условные отклонения в столбце 3 получены по формуле: $a = (x - A)/c$, где условная средняя ряда, имеющая максимальную частоту ($f = 171$), равна $A = 79$, а классовый интервал $c = 14$. Практически никаких вычислений для получения a не делают. Если ряд x возрастает сверху вниз с одинаковым интервалом (как в табл. 1.20), то против выбранного A записывают 0, а затем от нуля до конца ряда вверх с минусом, а вниз с плюсом пишут натуральные ряды чисел: 1, 2, 3, 4 и т. д. Если ряд x убывает сверху вниз, то в верхней половине, начиная от центрального нуля отклонения a будут со знаком плюс, а в нижней половине — со знаком минус.

3. Произведения частот на факториалы отклонений записаны в столбцах 4—7 табл. 1.20, где: $a_1 = a$, $a_2 = a(a - 1)$, $a_3 = a(a - 1)(a - 2)$, $a_4 = a(a - 1)(a - 2)(a - 3)$. Произведения указанного вида называются факториалами, откуда и получили название факториальные моменты. Таким образом, произведения в столбце fa_1 сверху вниз равны: $2 \cdot (-3) = -6$, $29 \cdot (-2) = -58$ и т. д.; произведения fa_2 равны: $2 \cdot (-3) \cdot (-4) = 24$, $29 \cdot (-2) \cdot (-3) = 174$ и т. д.; произведения fa_3 : $2 \cdot (-3) \cdot (-4) \cdot (-5) = -120$, $29 \cdot (-2) \cdot (-3) \cdot (-4) = -696$ и т. д.; произведения fa_4 : $2 \cdot (-3) \cdot (-4) \cdot (-5) \cdot (-6) = 720$; $29 \cdot (-2) \cdot (-3) \cdot (-4) \cdot (-5) = 3480$ и т. д.

4. Суммируем числа в столбцах 4—7 с учетом их знаков:

$$s_1 = 70, \quad s_2 = 616, \quad s_3 = -1434, \quad s_4 = 7056.$$

5. Факториальные моменты равны

$$w_1 = s_1/N, \quad w_2 = s_2/N, \quad w_3 = s_3/N, \quad w_4 = s_4/N, \quad (1.66)$$

где w_{1-4} — факториальные моменты; s_{1-4} — суммы произведений частот на факториалы отклонений; N — объем совокупности.

$$w_1 = 70/500 = 0,1400; \quad w_2 = 616/500 = 1,232; \\ w_3 = -1434/500 = -2,868; \quad w_4 = 7056/500 = 14,112.$$

6. На основании вычисленных значений факториальных моментов (w_{1-4}) находим условные моменты (m_{1-4}) по формулам:

$$m_1 = w_1, \quad m_2 = w_2 + w_1, \quad m_3 = w_3 + 3w_2 + w_1, \\ m_4 = w_4 + 6w_3 + 7w_2 + w_1. \quad (1.67)$$

Откуда:

$$m_1 = 0,14; \quad m_2 = 1,232 + 0,14 = 1,372; \\ m_3 = -2,868 + 3 \cdot 1,232 + 0,14 = 0,968; \\ m_4 = 14,112 + 6 \cdot (-2,868) + 7 \cdot 1,232 + 0,14 = 5,668.$$

Значения условных моментов, найденные здесь, полностью совпадают с вычисленными другим способом в § 1.22. Следовательно, подставив их в формулы (1.56)—(1.58), должны получить те же, что и в § 1.22, значения центральных моментов, измеренные в единицах классowego интервала.

б. Способ полного суммирования.

При отсутствии счетной автоматической машины условные моменты посредством факториальных при помощи простейших счетных средств вычислим способом сумм. На примере того же ряда распределения рассмотрим последовательность расчетов, в которых объединено вычисление факториальных и условных моментов (табл. 1.21).

1. Суммируем частоты f снизу вверх и записываем накопленные частоты в столбце f_1 : $16 + 42 = 58$, $58 + 121 = 179$, $179 + 171 = 350$ и т. д. Нижнюю частоту 16 повторяем во всех столбцах f_{1-4} в той же строке. Числа столбцов f_2, f_3, f_4 получаем точно также на основании чисел предыдущего столбца: $16 + 58 = 74$, $74 + 179 = 253$ и т. д.

Таблица 1.21. К вычислению условных моментов посредством факториальных по способу полного суммирования

	f	f_1	f_2	f_3	f_4
37	2	500	2070	5658	12 659
51	29	498	1570	3588	7 001
65	119	469	1072	2018	3 413
79	171	350	603	946	1 395
93	121	179	253	343	449
107	42	58	74	90	106
121	16	16	16	16	16
$c = 14$	$N = 500$	$s_1 = 2070$	$s_2 = 5658$	$s_3 = 12 659$	$s_4 = 25 039$

2. Суммируем числа во всех четырех столбцах и проверяем вычисление накопленных частот: верхняя частота столбца f_1 должна равняться $N = 500$; сумма каждого предыдущего столбца должна совпадать с верхней частотой последующего столбца. Вычисление и суммирование частот последнего столбца (f_4) необходимо повторить два раза.

3. Вычисляем условные моменты по формулам:

$$\begin{aligned} m_1 &= \frac{s_1}{N}, \quad m_2 = \frac{2s_2}{N} - m_1, \\ m_3 &= 2 \left[3 \left(\frac{s_3}{N} - 0,5m_2 \right) - m_1 \right], \\ m_4 &= 6 \left(\frac{4s_4}{N} - m_1 - m_3 \right) - 11m_2, \end{aligned} \quad (1.68)$$

откуда:

$$\begin{aligned} m_1 &= \frac{2070}{500} = 4,14; \quad m_2 = \frac{2 \cdot 5658}{500} - 4,14 = 18,492; \\ m_3 &= 2 \left[3 \left(\frac{12659}{500} - 0,5 \cdot 18,492 \right) - 4,14 \right] = 88,152; \\ m_4 &= 6 \left(\frac{4 \cdot 25039}{500} - 4,14 - 88,152 \right) - 11 \cdot 18,492 = 444,708. \end{aligned}$$

Несмотря на то, что величины условных моментов отличаются от таковых, полученных по предыдущему способу, подставив их в формулы (1.56)–(1.58), получим практически те же, что и в § 1.22 значения центральных моментов в единицах классového интервала:

$$\begin{aligned} \mu_2 &= m_2 - m_1^2 = 18,492 - 4,14^2 = 1,3524; \\ \mu_3 &= m_3 + m_1 (2m_1^2 - 3m_2) = 88,152 + 4,14 (2 \cdot 4,14^2 - 3 \cdot 18,492) = 0,3972; \\ \mu_4 &= m_4 + m_1 [3m_1 (2m_2 - m_1^2) - 4m_3] = 444,708 + 4,14 [3 \cdot 4,14 (2 \cdot 18,492 - 4,14^2) - 4 \cdot 88,152] = 5,2859. \end{aligned}$$

Для проверки вычисления второго центрального момента в единицах классového интервала служит формула

$$\mu_2 = \frac{2s_2 - s_1 \left(1 + \frac{s_1}{N} \right)}{N}, \quad (1.69)$$

где s_1, s_2 — суммы столбцов f_1 и f_2 из табл. 1.21; N — объем совокупности.

Если требуется вычислить среднее квадратическое отклонение, как обычно в именованных числах, то в формулу (1.69) вводится дополнительный множитель:

$$\sigma^2 = \frac{c^2}{N-1} \left[2s_2 - s_1 \left(1 + \frac{s_1}{N} \right) \right], \quad \sigma = \sqrt{\sigma^2}. \quad (1.69a)$$

По данным нашего примера (табл. 1.21)

$$\mu_2 = \frac{2 \cdot 5658 - 2070 \left(1 + \frac{2070}{500}\right)}{500} = 1,3524;$$
$$\sigma^2 = \frac{14^2}{500 - 1} \left[2 \cdot 5658 - 2070 \left(1 + \frac{2070}{500}\right) \right] = 265,61;$$
$$\sigma = \sqrt{265,61} = 16,3.$$

В § 1.22 по тем же данным $\sigma^2 = 265,07$ за счет того, что там не вычиталась единица из N в знаменателе, что допустимо при достаточно большом объеме выборки.

Так называемую поправку Бесселя $\frac{N}{N-1}$ можно вводить, когда $N < 100$, при вычислении сигмы извлечением корня из второго центрального момента. Если в знаменателе любой формулы вычисления сигмы или дисперсии стоит $N - 1$, это означает, что поправка Бесселя в нее введена и оценка дисперсии при этом стала несмещенной. В большинство этих формул указанная поправка входит.

Средняя арифметическая вычисляется по формуле:

$$M = x_1 + c (m_1 - 1), \quad (1.70)$$

где M — средняя арифметическая; x_1 — начальная варианта ряда; c — классовый интервал; m_1 — первый факториальный или первый условный момент (они равны между собой).

По данным нашего примера:

$$M = 37 + 14 (4,14 - 1) = 80,96 = 81,0.$$

Заметим, что для вычисления среднего квадратического отклонения и средней арифметической по способу «б» в табл. 1.21 кроме исходных данных достаточно получить частоты и суммы s_1 и s_2 лишь двух столбцов f_1 и f_2 .

На основе центральных моментов, дисперсии и средней можно рассчитать далее другие статистические показатели данного ряда (см. § 1.22).

в. Способ неполного суммирования.

1. По тем же данным составляем табл. 1.22, в которой варианты x возрастают сверху вниз.

2. Суммирование накопленных частот в столбцах f_{1-4} ведется снизу вверх, в каждом столбце оно заканчивается, на одну строку раньше не доходя до конца, по сравнению с предыдущим столбцом, как показано в таблице. Столбец f_1 : $16 + 42 = 58$, $58 + 121 = 179$ и т. д. Столбец f_2 : $16 + 58 = 74$, $74 + 179 = 253$, $253 + 350 = 603$, $603 + 469 = 1072$, $1072 + 498 = 1570$. Здесь суммирование должно было бы закончиться. Но для проверки вычислений получим еще одну накопленную частоту $1570 + 500 = 2070$, и в знак того, что она не будет участвовать в дальнейших расчетах, возьмем ее в скобки. Также сделаем и с конечными частотами f_3 и f_4 . Проверка заключается в том, что верхнее число последующего столбца должно равняться

Таблица 1.22. К вычислению факториальных моментов способом неполного суммирования

	f	f_1	f_2	f_3	
37	2	500	(2070)	—	—
51	29	498	1570	(3588)	—
65	119	469	1072	2018	(3413)
79	171	350	603	946	1395
93	121	179	253	343	449
107	42	58	74	90	106
121	16	16	16	16	16
<hr/>					
	$N = 500$	$s_1 = 2070$	$s_2 = 3588$	$s_3 = 3413$	$s_4 = 1966$

ся сумме частот предыдущего столбца у всех столбцов. Суммирование частот f_4 необходимо повторить для проверки.

3. Факториальные моменты вычисляем по формулам:

$$w_1 = \frac{s_1}{N}, \quad w_2 = \frac{2s_2}{N}, \quad w_3 = \frac{6s_3}{N}, \quad w_4 = \frac{24s_4}{N}, \quad (1.71)$$

где w_{1-4} — факториальные моменты; s_{1-4} — суммы частот в столбцах f_{1-4} табл. 1.22; N — объем совокупности.

$$w_1 = \frac{2070}{500} = 4,14; \quad w_2 = \frac{2 \cdot 3588}{500} = 14,352;$$

$$w_3 = \frac{6 \cdot 3413}{500} = 40,956; \quad w_4 = \frac{24 \cdot 1966}{500} = 94,368.$$

4. Условные моменты вычислим по формулам (1.67):

$$m_1 = w_1 = 4,14; \quad m_2 = 14,352 + 4,14 = 18,492;$$

$$m_3 = 40,956 + 3 \cdot 14,352 + 4,14 = 88,152;$$

$$m_4 = 94,368 + 6 \cdot 40,956 + 7 \cdot 14,352 + 4,14 = 444,708.$$

5. Центральные моменты вычислим по формулам (1.56)—(1.58):

$$\mu_2 = 18,492 - 4,14^2 = 1,3524;$$

$$\mu_3 = 88,152 + 4,14 (2 \cdot 4,14^2 - 3 \cdot 18,492) = 0,3972;$$

$$\mu_4 = 444,708 + 4,14 [3 \cdot 4,14 (2 \cdot 18,492 - 4,14^2) - 4 \cdot 88,152] = 5,2859.$$

6. Среднее квадратическое отклонение равно корню из второго момента, помноженному на классовой интервал:

$$\sigma = c \sqrt{\mu_2} = 14 \sqrt{1,3524} = 16,3.$$

Средняя арифметическая по формуле (1.70):

$$M = 37 + 14 (4,14 - 1) = 81,0.$$

Первые два центральных момента (μ_2, μ_3) возможно вычислить также непосредственно, используя суммы s_{1-3} , полученные в

табл. 1.22, по формулам:

$$\begin{aligned}\frac{\mu_2}{c^2} &= \frac{1}{N} \left[2s_2 + s_1 \left(1 - \frac{s_1}{N} \right) \right], \\ \frac{\mu_3}{c^3} &= \frac{1}{N} \left\{ 6(s_3 + s_2) + s_1 \left[1 + \frac{s_1 \left(\frac{2s_1}{N} - 3 \right) - 6s_2}{N} \right] \right\},\end{aligned}\quad (1.72)$$

где N — объем совокупности.

По данным нашего примера:

$$\begin{aligned}\frac{\mu_2}{c^2} &= \frac{1}{500} \left[2 \cdot 3588 + 2070 \left(1 - \frac{2070}{500} \right) \right] = 1,3524; \\ \frac{\mu_3}{c^3} &= \frac{1}{500} \left\{ 6(3413 + 3588) + \right. \\ &\quad \left. + 2070 \left[1 + \frac{2070(2 \cdot 2070/500 - 3) - 6 \cdot 3588}{500} \right] \right\} = 0,3972.\end{aligned}$$

Вычисление четвертого центрального момента непосредственно через суммы s_{1-4} нецелесообразно ввиду того, что соответствующая формула становится слишком громоздкой. Расчет этого момента, а при необходимости и остальных, можно выполнить или проверить по формулам:

$$\begin{aligned}\mu_2 &= w_2 + w_1(1 - w_1); \\ \mu_3 &= w_3 + 2(1,5w_2 + w_1[0,5 - 1,5w_2 - w_1(1,5 - w_1)]); \\ \mu_4 &= w_4 + 4[1,5w_3 + 1,75w_2 - w_1(w_3 + 3w_2 - 0,25 - \\ &\quad - w_1[1,5w_2 - 1 + 0,75w_1(2 - w_1)])],\end{aligned}\quad (1.73)$$

где w_{1-4} — факториальные моменты, вычисленные выше по способу «В».

Подставляя их значения в формулы (1.73), получим:

$$\begin{aligned}\mu_2/c^2 &= 14,352 + 4,14(1 - 4,14) = 1,3524; \\ \mu_3/c^3 &= 40,956 + 2(1,5 \cdot 14,352 + 4,14[0,5 - 1,5 \cdot 14,352 - \\ &\quad - 4,14(1,5 - 4,14)]) = 0,3974; \\ \mu_4/c^4 &= 94,368 + 4[1,5 \cdot 40,956 + 1,75 \cdot 14,352 - 4,14(40,956 + \\ &\quad + 3 \cdot 14,352 - 0,25 - 4,14[1,5 \cdot 14,352 - 1 + 0,75 \cdot 4,14(2 - \\ &\quad - 4,14)])] = 5,2852.\end{aligned}$$

Четвертый центральный момент можно вычислить также и по формуле:

$$\begin{aligned}\mu_4/c^4 &= 24[s_4 + 1,5(s_3 + 0,38889 \cdot s_2)]N^{-1} + w_1(1 - [4\mu_3 + \\ &\quad + w_1(6\mu_2 + w_1^2)]); \\ \mu_4/c^4 &= 24[1966 + 1,5(3413 + 0,38889 \cdot 3588)] \cdot 500^{-1} + 4,14(1 - \\ &\quad - [4 \cdot 0,3972 + 4,14(6 \cdot 1,3524 + 4,14^2)]) = 5,285.\end{aligned}\quad (1.74)$$

Проверку по формулам (1.73) и (1.74) следует проводить, однако, с несколько бóльшим числом значащих цифр в промежуточных вычислениях.

Таблица 1.23. К вычислению условных моментов посредством факториальных по способу суммирования от центра ряда

		f_1	f_2	f_3	f_4
37	2	2	2	2	—
51	29	31	33	—	—
65	119	150	—	—	—
79	$A = 171$	—	—	—	—
93	121	179	—	—	—
107	42	58	74	—	—
121	16	16	16	16	—
<hr/>					
$k = 7$	$N = 500$	$b = 183$	35	2	0
		$a = 253$	90	16	0
		$d = 70$	55	14	0
		$s = 436$	125	18	0

г. Способ суммирования от центра ряда.

1. Используя те же данные, составляем табл. 1.23, в которой против центральной варианты с максимальной частотой ($A = 171$) проводим черту, от нее вверх и вниз проводим еще по три черты так, как указано в таблице, т. е. каждый раз отступая на одну клетку от центра ряда.

2. Суммирование накопленных частот делается от центра ряда вверх и вниз. Сумма центральных вариантов: $171 + 150 + 179 = 500$, должна равняться объему совокупности.

Правильность суммирования накопленных частот в остальных столбцах проверяется по равенству сумм крайних (в центральном пробеле) вариант соответствующей сумме a или b . Например, $179 + 74 = a_1 = 253$, $74 + 16 = a_2 = 90$, $16 + 0 = a_3 = 16$, $150 + 33 = b_1 = 183$, $33 + 2 = b_2 = 35$, $2 + 0 = b_3 = 2$.

3. Верхнюю и нижнюю половины рядов чисел в столбцах f_{1-4} складываем отдельно, соответственно получим суммы b_{1-4} и a_{1-4} . Затем ниже записываем разности $d = a - b$ и суммы $s = a + b$ этих сумм.

4. Подставляя эти разности и суммы в формулы, в которых условные моменты выражены через факториальные

$$m_1 = d_1/N, \quad m_2 = (s_1 + 2s_2)/N, \quad m_3 = [d_1 + 6(d_2 + d_3)]/N, \quad (1.75)$$

$$m_4 = \{s_1 + 24[0,5(1,1667s_2 + 3s_3) + s_4]\} N^{-1},$$

получим:

$$m_1 = 70/500 = 0,14; \quad m_2 = (436 + 2 \cdot 125)/500 = 1,372;$$

$$m_3 = [70 + 6(55 + 14)]/500 = 0,968;$$

$$m_4 = \{436 + 24[0,5(1,1667 \cdot 125 + 3 \cdot 18) + 0]\} 500^{-1} = 5,668.$$

Таблица 1.24. К вычислению моментов по способу полуцентрального суммирования

	f	f_1	f_2	f_3	f_4
37	2	2	2	2	2
51	29	31	33	35	37
65	119	150	183	218	255
79	171	—	—	—	—
93	121	179	—	—	—
107	42	58	74	—	—
121	16	16	16	16	—
<hr/>					
$N = 500$		b 183	218	255	294
		a 253	90	16	0
		d 70	—128	—239	—294
		s 436	308	271	294

Значения условных моментов совпадают с найденными ранее; следовательно, по ним можно получить те же, что и раньше величины центральных моментов по формулам (1.56)—(1.58).

д. Способ полуцентрального суммирования.

1. Составим табл. 1.24, в первых двух столбцах которой приведены те же данные (x, f) , уже использованные для иллюстрации расчетов (разделы а—г).

2. В верхней половине таблицы образование накопленных частот ведем сверху вниз до черты против центральной варианты ($A = 171$): $2 + 29 = 31$, $31 + 119 = 150$, $2 + 31 = 33$, $33 + 150 = 183$, $2 + 33 = 35$ и т. д., в нижней половине таблицы накопленные частоты получаем суммированием снизу вверх: $16 + 42 = 58$, $58 + 121 = 179$, $16 + 58 = 74$ и т. д. до черточек, проставленных в таблице. Сумма центральных вариантов должна равняться объему совокупности $150 + 171 + 179 = N = 500$; нижняя частота в верхней половине таблицы должна равняться сумме b предыдущего столбца (например, $183 = b_1 = 183$); сумма любых двух верхних частот в нижней половине таблицы должна быть равна числу a (например, $179 + 74 = a_1 = 253$, $74 + 16 = a_2 = 90$).

3. Получаем суммы чисел верхней половины (b) и нижней половины (a) таблицы, затем записываем их разности $d = a - b$ и суммы $s = a + b$.

4. Факториальные моменты вычисляем по формулам:

$$w_1 = d_1/N; \quad w_2 = 2s_2/N; \quad w_3 = 6d_3/N; \quad w_4 = 24s_4/N, \quad (1.76)$$

где w_{1-4} — факториальные моменты; d и s — разности и суммы из табл. 1.24; N — объем совокупности.

По формулам (1.76):

$$w_1 = 70/500 = 0,14; \quad w_2 = 2 \cdot 308/500 = 1,232;$$

$$w_3 = 6 \cdot (-239)/500 = -2,868; \quad w_4 = 24 \cdot 294/500 = 14,112,$$

откуда условные моменты по формулам (1.67) равны:

$$\begin{aligned}m_1 &= w_1 = 0,14; & m_2 &= 1,232 + 0,14 = 1,372; \\m_3 &= -2,868 + 3 \cdot 1,232 + 0,14 = 0,968; \\m_4 &= 14,112 + 6 \cdot (-2,868) + 7 \cdot 1,232 + 0,14 = 5,668,\end{aligned}$$

что полностью совпадает со значениями, полученными другими способами. Следовательно, подставив их в формулы (1.56)—(1.58), мы вправе ожидать верных результатов при вычислении центральных моментов.

Симметричные ряды распределений для получения указанных моментов можно обработать любым из пяти приведенных способов. Для асимметричных рядов удобнее применять способы «б» или «в». Пример расчета по способу «б» дан в § 1.34 (показательное распределение).

§ 1.19. Асимметрия

Асимметрией, или скошенностью, называется мера отклонения распределения частот от симметричного их распределения относительно максимальной ординаты. Степень асимметричности распределения можно измерить двумя способами: через разность между средней арифметической и модой по формуле (1.77) и через отношение момента третьего порядка к кубу среднего квадратического отклонения по формуле (1.79).

Первый способ

$$K_{As} = (M - M_o)/\sigma, \quad (1.77)$$

где K_{As} — коэффициент асимметрии; M — средняя арифметическая; M_o — мода; σ — среднее квадратическое отклонение.

Подставив в формулу (1.77) соответствующие параметры ряда из табл. 1.03, получим:

$$K_{As} = (75,09 - 78,27)/5,99 = -0,53.$$

Отрицательное значение коэффициента соответствует отрицательной асимметрии, а положительное — положительной асимметрии (рис. 2, А).

Иногда в формуле (1.77) моду заменяют медианой, что также может характеризовать асимметрию:

$$K_{As} = \frac{M - M_e}{\sigma}. \quad (1.78)$$

Второй способ

$$A_s = \frac{\mu_3}{\sigma^3}, \text{ или } A_s = r_3, \quad (1.79)$$

где A_s — показатель асимметрии; μ_3 — центральный момент третьего порядка; σ — среднее квадратическое отклонение; r_3 — основной момент третьего порядка (табл. 1.18).

Подставляя μ_3 , вычисленный по данным табл. 1.19, и $\sigma = 6,056$

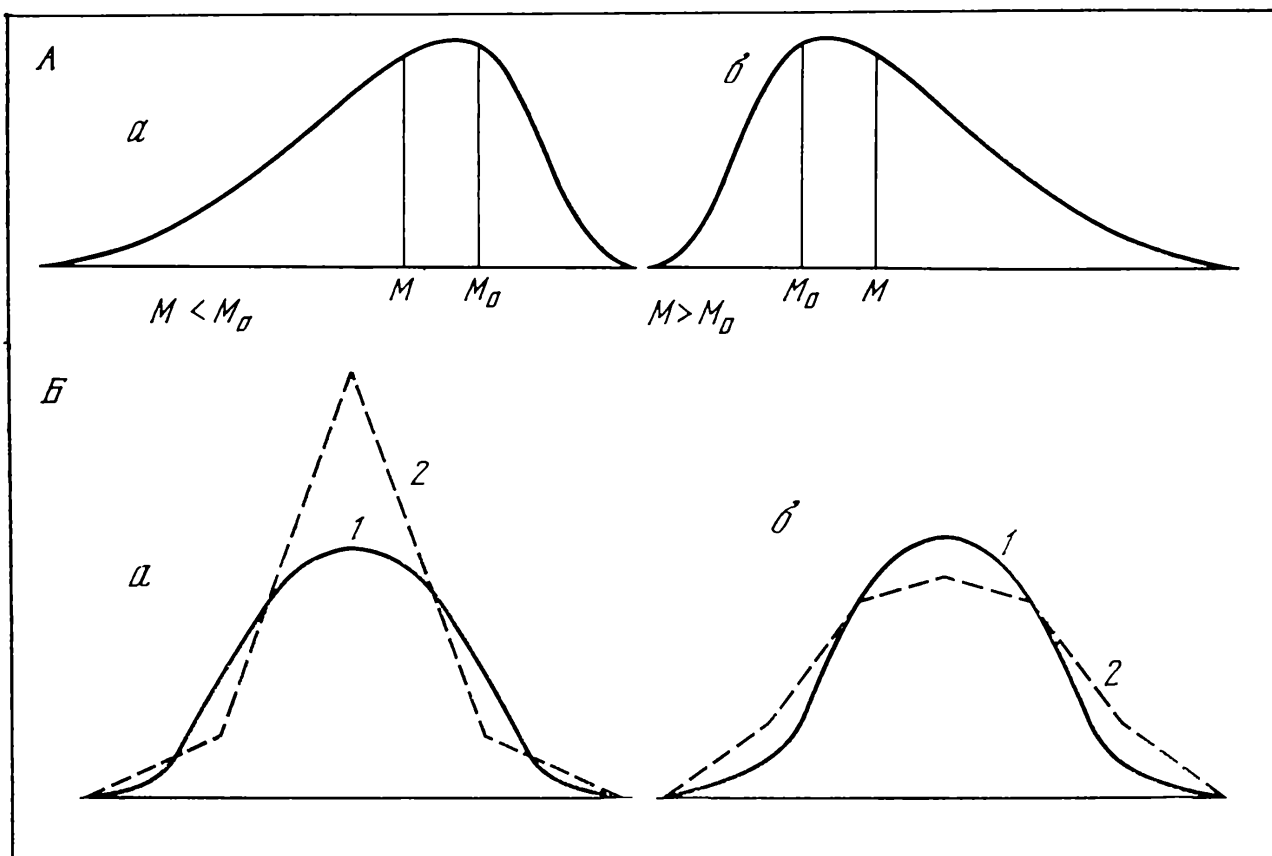


Рис. 2. Асимметрия (А: а — отрицательная, б — положительная) и эксцесс (Б: а — положительный, б — отрицательный)

1 — теоретическая и 2 — эмпирическая кривые распределения

в формулу (1.79), получим:

$$A_s = -\frac{53,036}{6,056^3} = -0,24.$$

Показатель асимметрии также имеет отрицательное значение, следовательно, распределение вариантов вегетационного периода сортов ячменя имеет отрицательную асимметрию.

Ошибка показателя асимметрии вычисляется по формуле:

$$m_A = \sqrt{\frac{6}{N}}, \quad (1.80)$$

где m_A — ошибка показателя асимметрии; N — объем выборки.

Для рассматриваемого примера о вегетационном периоде сортов ячменя:

$$m_A = \sqrt{\frac{6}{214}} = 0,17.$$

Следовательно:

$$A_s = -0,24 \pm 0,17.$$

Более точно ошибка асимметрии может быть вычислена по формуле:

$$m_A = \sqrt{\frac{6}{N+3}}, \quad (1.81)$$

которую следует применять для сравнительно небольших рядов.

Оценка достоверности показателя асимметрии производится по формуле (1.10) и табл. 3П. Число степеней свободы при этом принимается равным бесконечности ($\nu = \infty$).

$$t = \frac{A_s}{m_A},$$

где t — критерий Стьюдента (см. табл. 3П); A_s — показатель асимметрии; m_A — ошибка показателя асимметрии.

Для только что рассмотренного примера:

$$t = \frac{0,24}{0,17} = 1,41.$$

По табл. 3П при $\nu = \infty$ требуется величина критерия не меньше 1,96 для того, чтобы можно было говорить о существенности асимметрии. Следовательно, рассматриваемое распределение сортов ячменя по вегетационному периоду достаточно близко к симметричному.

В случаях, требующих повышенной точности, показатель асимметрии и его ошибку вычисляют с учетом несмещенности их величин по формулам:

$$A_s = \frac{r_3 \sqrt{N(N-1)}}{N-2}, \quad (1.82)$$

$$m_A = \sqrt{\frac{6N(N-1)}{(N-2)(N+1)(N+3)}}, \quad (1.83)$$

где обозначения те же, что и к формулам (1.79), (1.80). Для рассматриваемого примера:

$$A_s = \frac{-0,24 \sqrt{214(214-1)}}{214-2} = -0,2412;$$

$$m_A = \sqrt{\frac{6 \cdot 214(214-1)}{(214-2)(214+1)(214+3)}} = 0,166.$$

Минимальное число вариантов N_A , которое обеспечивает достоверность величины показателя асимметрии A_s , равно:

$$N_A = 54/A_s^2, \quad (1.83a)$$

В нашем примере: $N_A = 54/(-0,24)^2 = 54/0,0576 = 937$. Таким образом, для того чтобы надежно выявить особенности асимметрии ряда распределения величин вегетационного периода у ячменя необходимо взять 937 сортов или различных образцов этой культуры.

§ 1.20. Экссесс

Эмпирические распределения по сравнению с нормальным могут быть туповершинными или островершинными (см. рис. 2,Б). Степень отклонения эмпирической кривой распределения от нормальной теоретической кривой на своей вершине количественно выра-

жается показателем эксцесса по формуле:

$$E = \frac{\mu_4}{\sigma^4} - 3 \quad \text{или} \quad E = r_4 - 3, \quad (1.84)$$

где E — показатель эксцесса; μ_4 — центральный момент четвертого порядка (о вычислении его см. § 1.18); σ^4 — среднее квадратическое отклонение в четвертой степени ($\sigma^4 = \mu_2^2$); r_4 — основной момент четвертого порядка (см. § 1.18).

Эксцесс эмпирической кривой распределения сортов ячменя по продолжительности вегетационного периода определим, подставив в формулу (1.84) найденные в § 1.18 значения μ_4 и μ_2 :

$$E = \frac{3519,9473}{36,6801^2} - 3 = -0,38.$$

Ошибка показателя эксцесса равна удвоенной ошибке показателя асимметрии:

$$m_E = 2 \sqrt{\frac{6}{N}}, \quad (1.85)$$

где m_E — ошибка показателя эксцесса; N — объем выборки; $\sqrt{\frac{6}{N}}$ — ошибка показателя асимметрии.

$$m_E = 2 \sqrt{\frac{6}{214}} = \pm 0,34.$$

При вычислении m_E особенно для сравнительно небольших рядов применяют и более точную формулу:

$$m_E = \sqrt{\frac{24}{N+5}}, \quad (1.86)$$

где обозначения те же.

В наиболее ответственных случаях точные, несмещенные оценки показателя эксцесса и его ошибки вычисляют по формулам:

$$E = \frac{N^2 - 1}{(N-2)(N-3)} \left((r_4 - 3) + \frac{6}{N+1} \right), \quad (1.87)$$

$$m_E = \sqrt{\frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)}}, \quad (1.88)$$

где обозначения те же, что и к формулам (1.84), (1.85).

Для рассматриваемого примера:

$$E = \frac{214^2 - 1}{(214-2)(214-3)} \left((2,62 - 3) + \frac{6}{214+1} \right) = -0,389;$$

$$m_E = \sqrt{\frac{24 \cdot 214 \cdot (214-1)^2}{(214-3)(214-2)(214+3)(214+5)}} = 0,331.$$

Оценка достоверности показателя эксцесса производится по формуле (1.10):

$$t = \frac{E}{m_E}$$

где t — критерий Стьюдента (см. табл. 3П), в данном случае принимается число степеней свободы: $\nu = \infty$; E — показатель эксцесса; m_E — ошибка показателя эксцесса.

В рассмотренном здесь примере:

$$t = 0,38/0,34 = 1,12.$$

По табл. 3П при $\nu = \infty$ требуется, чтобы величина критерия была больше 1,96 для того, чтобы можно было считать распределение сортов ячменя по вегетационному периоду существенно отклоняющимся от нормального по своему эксцессу.

Показатель отрицательного эксцесса не может быть меньше, чем -2 , и это его значение указывает на то, что данная выборка состоит из вариантов, относящихся к различным совокупностям. «Провал» на вершине кривой при этом достигает оси абсцисс, и на графике получаются две различные кривые распределения. Положительный эксцесс по своей величине теоретически не ограничен. Эксцесс считается незначительным, если $|E| < 0,4$; чем меньше $|E|$, тем ближе распределение к нормальному.

Минимальное число вариантов (N_E), необходимое для достоверности показателя эксцесса E :

$$N_E = 216/E^2. \quad (1.88a)$$

В нашем примере: $N_E = 216/(-0,38)^2 = 216/0,1444 = 1496$. Следовательно, для получения достоверного представления об особенностях эксцесса распределения величин вегетационного периода ячменя необходимо составить совокупность из 1496 сортов и (или) различных образцов этой культуры.

§ 1.21. Энтропия

Многим биологическим явлениям, процессам, состояниям организмов и популяций в той или иной степени присуща неопределенность или различная информативность. Поэтому количественный показатель меры неопределенности, или энтропии, из теории информации должен найти широкое применение в биологических исследованиях. Энтропию возможно вычислить для любой системы, которая может принимать различные состояния с определенными вероятностями. В частности, посредством энтропийного анализа распределений фенодат в некоторой степени становится доступным объективное решение вопроса о том, какая из фенофаз содержит максимум полезной информации и, следовательно, должна наблюдаться более тщательно и глубоко, чем остальные. Если рассматривать комплекс фенодат по каждой фенофазе с точки зрения теории вероятностей как поле, образованное композицией независимых испытаний, что и представляет собой совокупность наблюдаемых видов растений, то к нему применимо понятие энтропии. Совокупности фенодат, состоящие из независимых испытаний, в то же время, следовательно, не могут считаться временными, или динамическими, рядами.

Энтропия определяется вероятностями всех элементарных событий данного поля, она служит мерой его неопределенности и вычисляется по общей формуле:

$$H(p_1, p_2 \dots p_k) = - \sum_{i=1}^k p_i \log_2 p_i, \quad (1.89)$$

где p_1, p_2, \dots, p_k — все вероятности данного поля в целом, которые можно заменить частотами распределений; $\log_2 p$ — двоичный логарифм вероятности. Для всякого поля, связанного с неопределенностью результата, энтропия всегда положительна, поэтому перед знаком суммы в формуле (1.89) стоит минус. k — число значений, которое может принимать данная система, в частности, это число классов распределения вариационного ряда.

Энтропия обычно измеряется в битах, что связано с применением двоичных логарифмов для ее вычисления, но она может быть выражена и в нитах, единицах, связанных с применением натуральных логарифмов или в десятичных единицах, когда энтропия вычисляется с помощью десятичных логарифмов. Практически энтропию удобнее вычислять с помощью натуральных логарифмов, а затем, для перевода в биты, делить полученную величину на $\ln 2 = 0,69314718$. Если энтропия найдена с помощью десятичных логарифмов, то для перевода в биты ее надо делить на $\lg 2 = 0,30103$. Напомним, что перевод логарифма числа (x) с основанием (b) в логарифм с основанием (a) производят по известной из алгебры формуле:

$$\log_a x = \frac{\log_b x}{\log_b a}, \quad (1.90)$$

Алгоритм вычисления энтропии покажем на примере взвешенного ряда, который представляет собой распределение по времени начала цветения 431 вида древесных растений из арборетума Ботанического института АН СССР в Ленинграде. Время начала цветения указано в столбце 1 табл. 1.25 в днях от 1 марта при помощи

Таблица 1.25. Распределение видов по датам зацветания

Границы класса, в днях от 1 марта	Середина класса, в днях от 1 марта, x	Частота (число видов), f	Частости, $\frac{f}{N} = p$	$\ln p$	$-p \ln p$
32,5—51,4	42	5	0,0116	—4,46541	0,0518
51,5—70,4	61	38	0,0882	—2,42815	0,2142
70,5—89,4	80	120	0,2784	—1,27870	0,3560
89,5—108,4	99	160	0,3712	—0,99101	0,3679
108,5—127,4	118	68	0,1578	—1,84643	0,2914
127,5—146,4	137	20	0,0464	—3,07046	0,1425
146,5—165,4	156	12	0,0278	—3,58272	0,0996
165,5—184,4	175	6	0,0139	—4,27587	0,0594
184,5—203,4	194	1	0,0023	—6,07485	0,0140
203,5—222,5	213	1	0,0023	—6,07485	0,0140
		431	0,9999		1,6108

переводной табл. 24П. Для вычисления энтропии нужны только частоты ряда, которые путем деления на объем ряда (431) преобразуем в частоты (столбец 4). Натуральные логарифмы частот (столбец 5) умножаем на частоты (столбец 6) и суммируем, получаем энтропию в нитах $H_n = 1,6108$, делим ее на $\ln 2$, что дает энтропию в битах:

$$H_6 = 1,6108 / 0,69315 = 2,324.$$

Подобным образом были вычислены величины энтропии для рядов распределения и других фаз этого же массива древесных растений. По своей убывающей информативности фазы располагаются в следующем порядке (в скобках величина энтропии в нитах): начало цветения (1,611), конец цветения (1,729), начало распускания листьев (1,845), начало осенней раскраски листьев (1,895), созревание плодов (1,917) и начало осеннего листопада (1,917). Более других содержит информации распределение видов по началу цветения, мера неопределенности поля вероятностей которого наименьшая (1,611). У травянистых многолетников путем множественного корреляционного анализа, т. е. принципиально иными методами, установлено, что центром корреляционной плеяды их фаз в Москве является также начало цветения, а на втором месте по силе корреляционных взаимосвязей стоит конец цветения, что также полностью соответствует порядку возрастания величин энтропии этих фаз у деревьев и кустарников в Ленинграде. Следовательно, у травянистых и древесных многолетников основной из фаз, определяющей ход других, является начало цветения, которая и должна наблюдаться в первую очередь.

§ 1.22. Вычисление параметров взвешенного ряда

Приведенную здесь схему расчета следует применять при необходимости сделать сравнительно полный анализ изучаемой совокупности. При разработке алгоритма имелось в виду: свести к минимуму вычислительную работу и вместе с тем обеспечить требуемую точность и правильность расчетов путем введения проверочных операций. Методика подготовки выборки для обработки в виде взвешенного вариационного ряда приводится в § 1.01.

Вычислим параметры ряда распределения соцветий нивяника по диаметру (в мм) (табл. 1.26).

1. В столбце 2 выбираем варианту, имеющую наибольшую частоту. Такой здесь оказалась центральная варианта 79 с частотой 171. Следовательно, начало отсчета $A = 79$, а классовый интервал $c = 14$.

2. По формуле $a = \frac{x - A}{c} = \frac{x - 79}{14}$ кодируем варианты (столбец 4).

3. Частоты f перемножаем с 1—4 степенями условных отклонений a (столбцы 5—8).

4. Суммируем числа в столбцах 3, 5—8.

5. Прежде, чем переходить к вычислению условных моментов, проверим сделанные вычисления. Для этого переносим на один класс

Таблица 1.26. К вычислению моментов ряда распределения величин диаметра соцветия у нивяника

Границы класса	Середина класса, x	Частота, f		fa	fa^2	fa^3	fa^4	$a + 1$	$(a + 1)^4$	$(a + 1)^4 f$
30—43	37	2	—3	—6	18	—54	162	—2	16	32
44—57	51	29	—2	—58	116	—232	464	—1	1	29
58—71	65	119	—1	—119	119	—119	119	0	0	0
72—85	79	171	0	0	0	0	0	1	1	171
86—99	93	121	1	121	121	121	121	2	16	1936
100—113	107	42	2	84	168	336	672	3	81	3402
114—128	121	16	3	48	144	432	1296	4	256	4096
$c = 14$	$A = 79$	500	$k = 7$	70	686	484	2834			9666

вверх начало отсчета условных отклонений, т. е. кодируем ряд x по формуле: $1 + a = \frac{x - 65}{14}$ и записываем новые отклонения в столбце 9, затем их возводим в 4-ю степень (столбец 10) и перемножаем на частоты. Сумма чисел столбца 11 должна быть равна следующему выражению:

$$\Sigma f(a + 1)^4 = \Sigma fa^4 + 4\Sigma fa^3 + 6\Sigma fa^2 + 4\Sigma fa + N; \quad (1.91)$$

$$9666 = 2834 + 4 \cdot 484 + 6 \cdot 686 + 4 \cdot 70 + 500;$$

$$9666 = 9666. \text{ Вычисления сумм сделаны верно.}$$

Если начало отсчета отклонений перенести вниз от вначале принятого нуля, например, против варианты 93 с частотой 121, то проверку надо делать по формуле:

$$\Sigma f(a - 1)^4 = \Sigma fa^4 - 4\Sigma fa^3 + 6\Sigma fa^2 - 4\Sigma fa + N. \quad (1.92)$$

6. Находим условные моменты:

$$m_1 = \frac{\Sigma fa}{N} = \frac{70}{500} = 0,1400; \quad m_2 = \frac{\Sigma fa^2}{N} = \frac{686}{500} = 1,3720;$$

$$m_3 = \frac{\Sigma fa^3}{N} = \frac{484}{500} = 0,9680; \quad m_4 = \frac{\Sigma fa^4}{N} = \frac{2834}{500} = 5,6680.$$

Моменты рекомендуется вычислять не менее чем с четырьмя знаками после запятой.

7. Вычисляем по формулам (1.56)—(1.58) правую часть следующих уравнений:

$$\frac{\mu_2}{c^2} = m_2 - m_1^2 = 1,3720 - 0,1400^2 = 1,3524;$$

$$\begin{aligned} \frac{\mu_3}{c^3} &= m_3 + m_1(2m_1^2 - 3m_2) = 0,9680 + \\ &+ 0,1400(2 \cdot 0,1400^2 - 3 \cdot 1,3720) = 0,3973; \end{aligned}$$

$$\begin{aligned} \frac{\mu_4}{c^4} &= m_4 + m_1[3m_1(2m_2 - m_1^2) - 4m_3] = \\ &= 5,6680 + 0,1400[3 \cdot 0,1400(2 \cdot 1,3720 - 0,1400^2) - \\ &- 4 \cdot 0,9680] = 5,2861. \end{aligned}$$

8. Проверяем по формулам (1.59), (1.60) вычисление центральных моментов (деленных на степени классового интервала):

$$\frac{\mu_3}{c^3} = m_3 - m_1 \left(3 \frac{\mu_2}{c^2} + m_1^2 \right) = 0,9680 - \\ - 0,1400 (3 \cdot 1,3524 + 0,1400^2) = 0,3973;$$

$$\frac{\mu_4}{c^4} = m_4 - m_1 \left[\frac{4\mu_3}{c^3} + m_1 \left(\frac{6\mu_2}{c^2} + m_1^2 \right) \right] = \\ = 5,6680 - 0,1400 [4 \cdot 0,3973 + \\ + 0,1400 (6 \cdot 1,3524 + 0,1400^2)] = 5,2861.$$

Значения μ_3/c^3 и μ_4/c^4 практически совпадают с вычисленными выше, следовательно, расчеты произведены верно.

9. Находим центральные моменты из уравнений:

$$\mu_2/c^2 = 1,3524; \quad \mu_3/c^3 = 0,3973; \quad \mu_4/c^4 = 5,2861,$$

где $c^2 = 14^2 = 196$; $c^3 = 14^3 = 2744$; $c^4 = 14^4 = 38416$; (c — классовый интервал).

$$\mu_2 = 196 \cdot 1,3524 = 265,0704; \quad \mu_3 = 0,3973 \cdot 2744 = 1090,1912; \\ \mu_4 = 38416 \cdot 5,2861 = 203070,8176.$$

10. Средняя арифметическая (1.37):

$$M = A + m_1 c = 79 + 0,1400 \cdot 14 = 81,0.$$

11. Среднее квадратическое отклонение:

$$\sigma = \sqrt{\mu_2} = \sqrt{265,0704} = 16,28.$$

12. Коэффициент асимметрии (1.79):

$$A_s = \frac{\frac{\mu_3}{c^3}}{\frac{\mu_2}{c^2} \sqrt{\frac{\mu_2}{c^2}}} = \frac{0,3973}{1,3524 \sqrt{1,3524}} = 0,253.$$

13. Эксцесс (1.84):

$$E = \frac{\frac{\mu_4}{c^4}}{\left(\frac{\mu_2}{c^2} \right)^2} - 3 = \frac{5,2861}{1,3524^2} - 3 = -0,11.$$

14. Коэффициент вариации (1.43):

$$v = \frac{\sigma}{M} \cdot 100 = \frac{16,28}{80,96} \cdot 100 = 20,1\%.$$

Пользуясь полученными здесь данными, можно вычислить также другие параметры и кривые распределения взвешенного ряда в зависимости от цели обработки.

Кроме указанной в § 1.17 существует формула определения моды через показатели асимметрии (A_s), эксцесса (E), величины сигмы (σ)

и средней арифметической (M):

$$M_o = M - \frac{A_s \cdot \tau (E/6 + 1)}{2(5/6E - A_s^2 + 1)}, \quad (1.93)$$

после чего возможно приближенное вычисление медианы:

$$M_e = \frac{M_o + 2M}{3} \quad (1.94)$$

По данным рассматриваемого примера:

$$M_o = 81 - \frac{0,253 \cdot 16,28 (-0,11/6 + 1)}{2 \left(\frac{5 \cdot (-0,11)}{6} - 0,253^2 + 1 \right)} = 78,605.$$

$$M_e = \frac{78,605 + 2 \cdot 81}{3} = 80,205.$$

Формулы (1.93), (1.94) удобны при вычислениях на ЭВМ, когда желательно избежать промежуточного обращения к исходным данным, чтобы найти границы модального и медианного классов (§ 1.15, 1.17).

Рассмотренный пример относится к ряду, распределение частот которого довольно симметрично. Вычисление статистик взвешенного ряда с резко асимметричным распределением частот покажем на примере распределения величин коэффициента вариации 175 биологических признаков, для каждого из которых коэффициент вариации был получен в свою очередь в результате обработки взвешенного вариационного ряда достаточного объема. В табл. 1.27 приведены: x — величины коэффициента вариации от 10 до 190; f — частоты; a — условные отклонения от варианты $A = 30$, полученные по формуле: $(x - 30)/20$; fa , fa^2 , fa^3 , fa^4 — произведения степеней отклонений на частоты; $a - 1$, $(a - 1)^4$, $(a - 1)^4 f$ — проверочные вычис-

Таблица 1.27. К вычислению моментов ряда распределения величин коэффициента вариации биологических признаков

	f			fa^2	fa^3	fa^4	$a - 1$	$(a - 1)^4$	$(a - 1)^4 f$
10	89	-1	-89	89	-89	89	-2	16	1424
30	45	0	0	0	0	0	-1	1	45
50	19	1	19	19	19	19	0	0	0
70	9	2	18	36	72	144	1	1	9
90	3	3	9	27	81	243	2	16	48
110	1	4	4	16	64	256	3	81	81
130	3	5	15	75	375	1875	4	256	768
150	3	6	18	108	648	3888	5	625	1875
170	2	7	14	98	686	4802	6	1296	2592
190	1	8	8	64	512	4096	7	2401	2401
175			16	532	2368	15 412			9243

ления с нулем против варианты $A = 50$, число классов ряда $k = 10$, интервал ряда $c = 20$, объем ряда $N = 175$. Так как в этом примере для проверки вычислений число A перенесено не вверх, а вниз на один класс, то применяем формулу (1.92):

$$15412 - 4 \cdot 2368 + 6 \cdot 532 - 4 \cdot 16 + 175 = 9243,$$

что совпадает с итогом последнего столбца табл. 1.27; следовательно, расчеты в ней выполнены верно.

Условные моменты вычисляем по тем же формулам:

$$m_1 = 16/175 = 0,0914286; \quad m_2 = 532/175 = 3,04;$$

$$m_3 = 2368/175 = 13,53143; \quad m_4 = 15412/175 = 88,068576.$$

Средняя арифметическая (1.37):

$$M = A + m_1 c = 30 + 20 \cdot 0,0914286 = 31,828572.$$

Среднее квадратическое отклонение ввиду сравнительно небольшого объема ряда лучше определить по формуле:

$$\sigma = c \sqrt{\frac{\sum fa^2 - m_1 \sum fa}{N - 1}} \quad (1.95)$$

где обозначения те же, что и раньше.

$$\sigma = 20 \sqrt{\frac{532 - 0,0914286 \cdot 16}{175 - 1}} = 34,922,$$

что является несмещенной оценкой сигмы, в отличие от формулы $\sigma = \sqrt{\mu_2}$, применяемой для больших совокупностей, и где поэтому можно пренебречь несущественным смещением оценки сигмы. Дальнейшие расчеты других параметров ряда можно провести, пользуясь теми же формулами, по которым обработан симметричный ряд в первой половине данного параграфа.

§ 1.23. Доверительные интервалы статистических параметров

Ценность статистически обработанных данных заключается также и в том, что вычисленные средние, коэффициенты вариации, коэффициенты корреляции, критерии различия и другие показатели получают количественные границы доверия, которые обозначают возможные колебания показателя в меньшую и большую стороны в пределах доверительного интервала. В зависимости от типа распределения данных доверительный интервал рассчитывается двумя способами.

1. В симметричных распределениях, близких к нормальному, размах отклонений данных от средней арифметической обычно равен приблизительно 3σ в ту и другую стороны, поэтому считают, что все возможные отклонения параметра от его истинной величины будут лежать в пределах $\pm 3\sigma$. Отклонения от средней арифметиче-

ской для их сравнимости выражают в долях сигмы:

$$t = \frac{x - M}{\sigma},$$

где t — нормированное отклонение от средней арифметической; x — любая варианта совокупности; M — средняя арифметическая; σ — среднее квадратическое отклонение.

При расчете доверительного интервала для трех стандартных доверительных уровней: $P_1 = 95\%$; $P_2 = 99\%$; $P_3 = 99,9\%$, t выбирается по табл. 3П, по числу степеней свободы, соответствующему данному параметру.

Если объемы выборок достаточно велики, то принимаются t равные: 1,960; 2,576; 3,291, или округленно 2,0; 2,6; 3,3 по тем же доверительным уровням, т. е. берется последняя строка табл. 3П.

Доверительный интервал средней арифметической строится по формуле:

$$M - tm_M < \text{истинное значение} < M + tm_M. \quad (1.96)$$

Пример: Доверительный интервал средней арифметической из § 1.16 построим на 95% доверительном уровне, при числе степеней свободы $\nu = 14 - 1 = 13$, откуда по табл. 3П $t = 2,160$. Средняя равна $M = 74,9$, ее ошибка $m_M = 1,9$, следовательно:

$$74,9 - 2,160 \cdot 1,9 < \text{истинное значение} < 74,9 + 2,160 \cdot 1,9, \text{ т. е.}$$

истинное значение средней арифметической находится в пределах не меньше 70,796 и не больше 79,004:

$$70,8 < M < 79,0.$$

Подставив в формулу (1.96) вместо M и m_M величину любого параметра и его ошибки, например среднего квадратического отклонения, коэффициента вариации, коэффициента корреляции и других, можно построить для них доверительный интервал и, таким образом, оценить эти параметры.

2. Когда доверительный интервал необходимо построить при неизвестном типе распределения вариант, применяется неравенство Чебышева. В этом случае подразумевается, что размах отклонений от средней арифметической в совокупности с неизвестным распределением может быть больше, чем в нормальном распределении, т. е. больше, чем 3σ .

Согласно неравенству Чебышева истинная средняя арифметическая находится в интервале:

$$M - \frac{m_M}{\sqrt{1-P'}} \leq \text{истинное значение} \leq M + \frac{m_M}{\sqrt{1-P'}}, \quad (1.97)$$

где M — средняя арифметическая, рассчитанная по выборке; m_M — ее ошибка; P' — доверительный уровень в долях единицы или $1 - P' = W'$ — уровень значимости.

Построим доверительный интервал для средней из предыдущего примера по неравенству Чебышева: $M = 74,9$; $m_M = 1,9$; примем доверительный уровень $P_1 = 95\%$, т. е. уровень значимости

$W'_1 = 1 - P'_1 = 1 - 0,95 = 0,05$. По формуле (1.97):

$$74,9 - \frac{1,9}{\sqrt{0,05}} \leq \text{истинная средняя} \leq 74,9 + \frac{1,9}{\sqrt{0,05}}$$

Следовательно, согласно неравенству (1.97), средняя арифметическая с вероятностью 0,95 находится в пределах от 66,4 до 83,4, т. е. в более широком интервале, чем при нормальном распределении. При доверительном уровне $P_1 = 95\%$ истинная величина статистического параметра (например, средней арифметической) находится в пределах:

$$\text{от } M - 4,47m_M \text{ до } M + 4,47m_M,$$

(где $4,47 = \frac{1}{\sqrt{0,05}}$), каково бы ни было распределение данных в рассматриваемой выборке.

Если требуется определить доверительный интервал статистического параметра лишь с одной стороны (большей или меньшей), то значение t по сравнению с формулой (1.96) принимается несколько меньшим, по табл. 25П. Для трех стандартных уровней: $P_1 = 95\%$; $P_2 = 99\%$ и $P_3 = 99,9\%$ t при одностороннем доверительном интервале для больших ($N \geq 30$) выборок соответственно равно 1,645; 2,326 и 3,090. Используя пример из § 1.04, где средняя арифметическая равна $M = 75,05$, а ее ошибка $m_M = \pm 0,41$, вычислим доверительный интервал в большую сторону на 95% доверительном уровне:

$$\text{истинное значение} \leq M + tm_M; \quad (1.98)$$

$$M + tm_M = 75,05 + 1,645 \cdot 0,41 = 75,72.$$

Следовательно, правая или большая граница одностороннего доверительного интервала рассматриваемой средней арифметической составляет 75,72. Односторонний интервал с левой или меньшей стороны находится путем уменьшения M на величину tm_M :

$$M - tm_M \leq \text{истинное значение}, \quad (1.99)$$

где t принимается для больших выборок согласно выше приведенным значениям для одностороннего интервала или по табл. 25П, если выборки небольшие, при числе степеней свободы $N - 1$.

§ 1.24. Определение репрезентативности и методы отбора вариант выборок

Генеральной совокупностью называется вся совокупность особей или предметов изучаемого объекта. Если изучается средняя высота растений какого-либо сорта, то генеральной совокупностью являются все особи данного сорта, выращенные в сходных условиях. Большинство биометрических исследований проводится на некоторой части генеральной совокупности; эта часть называется выборкой, или выборочной совокупностью. Статистические показатели, полученные по выборке, распространяются затем на всю генеральную

совокупность данного объекта. Численность генеральной совокупности может быть известной, а чаще ее определить трудно. Объем выборки, достаточный для достоверности получаемой по ней средней арифметической, можно определить различными способами. Объем репрезентативной (т. е. достаточной по численности) выборки для того случая, когда известен объем генеральной совокупности, исчисляется по формуле

$$N = 1 / \left[\left(\frac{\Delta}{t\sigma} \right)^2 + \frac{1}{N_0} \right], \quad (1.100)$$

где N — искомый объем выборки; t — критерий достоверности Стьюдента; в зависимости от доверительного уровня исследования он равен: $t = 1,96$ при $P'_1 = 0,95$; $t = 2,58$ при $P'_2 = 0,99$ и $t = 3,30$ при $P'_3 = 0,999$; N_0 — численность генеральной совокупности; Δ — допустимая погрешность определения средней арифметической в именованных числах; σ — среднее квадратическое отклонение.

Если сигму нельзя вычислить по предыдущим опытам с данным объектом, ее можно определить по приближенным способам, рассмотренным в § 1.11.

Определим репрезентативность выборки для изучения вегетационного периода совокупности сортов ячменя. Предположим, что существует всего около 2000 сортов этой культуры. Из предыдущих расчетов известно, что в этой совокупности для выборки $N = 214$ среднее квадратическое отклонение $\sigma = 6,07$ дня.

Примем допустимую погрешность определения средней арифметической $\Delta = 1$ день, тогда по формуле (1.100) при $t = 1,96$:

$$N = 1 / \left[\left(\frac{1}{1,96 \cdot 6,07} \right)^2 + \frac{1}{2000} \right] = 132 \text{ сорта,}$$

Таким образом, для изучения вегетационного периода совокупности сортов ячменя достаточно было взять выборку в 132 сорта при условии их случайного отбора.

Репрезентативность выборки для достоверности ее средней арифметической можно определить, минуя расчет средней арифметической, сигмы и ошибки средней арифметической по формуле:

$$(N - 1) / [N \sum x^2 / (\sum x)^2 - 1] \leq 9, \quad (1.101)$$

где $\sum x$ — сумма вариантов ряда; N — число вариантов в ряду; $\sum x^2$ — сумма квадратов вариант ряда.

Если значение левой части неравенства (1.101) будет больше 9, то данный объем выборки достаточен для достоверности средней арифметической совокупности. Оценку значимости по формуле (1.101) можно делать также более точно по критерию Фишера при $v(1) = 1$ и $v(2) = N - 1$ (см. § 4.01 и табл. 9П).

При отборе нужного числа вариантов из генеральной совокупности необходимо прибегать к методу рендомизации, который обеспечивает случайность выбора вариантов. Например, для производства каких-либо наблюдений или измерений следует отбирать не «типичные» и не «средние» экземпляры, а те, которые попадают в выборку по схе-

ме отбора независимо от размещения и состояния растений или их органов. Применяют в основном метод механического отбора и жеребьевку. При механическом отборе отсчитывают каждое пятое, или десятое, или двадцатое и т. д. растение или рядок. При жеребьевке всем растениям делянки присваивают номера, затем записывают их на билеты и по жребью выбирают нужное число особей или отбирают их по таблице случайных чисел [см. (Янко, 1961, табл. 36)]. Вторым принципом отбора выборки — равномерность распределения отбираемых вариантов по делянке, опытному полю, по буртам, ящикам и т. д. Например, нельзя брать все 20 растений какой-либо выборки из одного места делянки или 20 плодов из одного ящика, если ящиков много и т. д. Можно отобрать 20 требуемых растений, например по диагонали делянки через равные промежутки так, чтобы отбираемые особи распределялись равномерно по всей длине диагонали.

При подсчете репрезентативности выборки, например, из совокупности сортов растений по какому-либо хозяйственно важному признаку, получаемое число может послужить при определенных условиях указанием на качественный состав данной совокупности сортов. Если, например, при объеме генеральной совокупности, равном 2000 сортов, получается при подсчете по формуле (1.100), что для достоверного изучения длительности вегетационного периода сортов ячменя достаточно располагать выборкой в 132 сорта, это помимо чисто статистического смысла данной величины означает также и хозяйственную ненужность остальных 1868 сортов, так как они повторяют по данному признаку уже имеющиеся сорта. Поэтому использование таких сортов может быть оправдано лишь в том случае, если они обладают другими ценными признаками, не встречающимися у остальных сортов. Следовательно, сортооценка любой совокупности сортов должна производиться по достаточно репрезентативному комплексу хозяйственно и биологически важных признаков.

В биологических исследованиях часто встречаются случаи, когда подлежащая изучению совокупность имеет неизвестную численность. Достаточный объем выборки из таких совокупностей определяют по формуле:

$$N = \left(\frac{t\sigma}{\Delta} \right)^2, \quad (1.102)$$

где N — искомый объем выборки; t — критерий достоверности [о его величине см. пояснение к формуле (1.100)]; σ — среднее квадратическое отклонение. Если сигма данного объекта не известна, ее можно приближенно определить по методам, рассмотренным в § 1.11; Δ — допустимая погрешность определения средней арифметической.

Определим достаточный объем выборки для примера, приведенного в начале параграфа, в предположении, что объем совокупности неизвестен, $t = 1,96$, $\sigma = 6,07$ дня, $\Delta = 1$ день. Тогда по формуле (1.102):

$$N = \left(\frac{1,96 \cdot 6,07}{1} \right)^2 = 142.$$

Формулы (1.100) и (1.102) дают только приближенные значения численности выборок, результаты, получаемые по ним, сильно зависят от принимаемой величины допустимой погрешности средней арифметической. Проверить правильность назначения достаточной численности выборки можно, вычисляя затем по ней показатель точности опыта (см. § 1.14), который должен быть не больше 5%.

Если возможно приблизительно определить величину варьирования вариант совокупности в процентах, то для нахождения достаточного объема выборки можно применить формулу:

$$N = \left(\frac{tv}{\varepsilon} \right)^2, \quad (1.103)$$

где N — искомый минимальный объем выборки; v — коэффициент вариации, в процентах; ε — допустимая погрешность, в процентах; t — критерий достоверности, см. пояснение к формуле (1.100).

Рекомендуется принимать ε равным 3% и не больше 5%. Если известны точно или приблизительно средняя арифметическая и сигма (см. § 1.11), то коэффициент вариации можно подсчитать (см. § 1.13), а затем подставить его значение в формулу (1.103). Достаточная численность выборки при $t = 1,96$, $v = 10\%$ и $\varepsilon = 3\%$, по формуле (1.103):

$$N = \left(\frac{1,96 \cdot 10}{3} \right)^2 = 43.$$

При исходных данных: $t = 1,96$; $\varepsilon = 3\%$ формула (1.103) получает вид:

$$N = 0,43v^2.$$

Минимальный (при $t = 1,96$ и $\varepsilon = 5\%$) и максимальный (при $t = 3,30$ и $\varepsilon = 1\%$) объемы достаточных выборок можно определить по формуле (1.103) в следующих ее вариантах.

Минимальная выборка:

$$N = 0,1537v^2.$$

Максимальная выборка:

$$N = 10,89v^2.$$

При $v = 10\%$ минимальная и максимальная выборки соответственно равны: 15 особей и 1089 особей. Размеры минимальных выборок в зависимости от коэффициента вариации исходных данных приведены в табл. 23П.

При альтернативном варьировании и известной численности генеральной совокупности изучаемого объекта объем выборки, достаточный для достоверности доли, можно приближенно определить по формуле (1.100):

$$N = 1 / \left[\left(\frac{\Delta}{\sigma t} \right)^2 + \frac{1}{N_0} \right]$$

Для часто применяемого уровня достоверности в 95%, при $t = 1,96$ и, принимая сигму $\sigma = 0,5$ (когда она неизвестна), формула (1.100) примет вид:

$$N = 1/[1,042\Delta^2 + 1/N_0], \quad (1.104)$$

где N — искомый объем выборки; N_0 — объем генеральной совокупности; Δ — допустимая погрешность при определении доли; σ — среднее квадратическое отклонение изучаемой совокупности; если оно неизвестно, следует принять максимальное его значение при альтернативном разделении дат $\sigma = 0,5$; t — показатель достоверности суждения, в зависимости от требуемой точности опыта он равен: 1,96; 2,58; 3,30.

Определим достаточный размер выборки, который требуется для достоверности определения доли корнеплодов моркови, поражаемых фомозом при механическом их повреждении (по данным из § 1.27). Сигма этой совокупности известна и равна: $\sigma = \sqrt{0,63 \cdot 0,37} = 0,48$, имеющаяся в наличии численность генеральной совокупности $N_0 \approx \approx 600$ корнеплодов. Примем $\Delta = 0,1$ и $t = 1,96$. По формуле (1.100):

$$N = 1 / \left[\left(\frac{0,1}{0,48 \cdot 1,96} \right)^2 + \frac{1}{600} \right] = 77.$$

Таким образом, для достоверного определения доли пораженных корнеплодов достаточно 77 корнеплодов.

По формуле (1.104) достаточная величина выборки для совокупности с неизвестной сигмой, которая принимается $\sigma = 0,5$ и при $\Delta = 0,1$; $N_0 = 600$:

$$N = 1/[1,042 \cdot 0,1^2 + 1/600] = 83.$$

Увеличение выборки на 6 корнеплодов по сравнению с предыдущим результатом произошло из-за увеличения сигмы с 0,48 до 0,50.

При планировании опыта или обследования необходимый объем выборки, обеспечивающий достоверность доли при альтернативном варьировании признака в зависимости от требуемой точности при неизвестном объеме генеральной совокупности, может быть приближенно определен по формуле (1.102):

$$N = \left(\frac{t\sigma}{\Delta} \right)^2,$$

где N — искомый объем выборки; t — показатель достоверности суждения, в зависимости от требуемой точности опыта равен: 1,96; 2,58; 3,30; σ — среднее квадратическое отклонение. Если сигма неизвестна, следует принять ее максимально возможное значение при альтернативном разделении вариантов равным 0,5; Δ — допустимая погрешность определения доли в частях единицы.

Для данных табл. 1.04 о доле мутантов среди особей ячменя по формуле (1.02) требуемый минимум объема выборки для достоверности доли мутантов, при $t = 1,96$, $\sigma = \sqrt{pq} = \sqrt{0,06 \cdot 0,94} = 0,24$

и $\Delta = 0,02$

$$N = \left(\frac{1,96 \cdot 0,24}{0,02} \right)^2 = 553.$$

Величина выборки, в которой гарантировано наличие хотя бы одной особи, обладающей изучаемым признаком, определяется по формуле:

$$N = \frac{\lg(1 - P'_{1-3})}{\lg(1 - p)}, \quad (1.105)$$

где N — объем выборки; $P'_1 = 0,95$; $P'_2 = 0,99$; $P'_3 = 0,999$; p — доля особей в генеральной совокупности, обладающих изучаемым признаком. При делении логарифмов их необходимо преобразовать из искусственной в естественную форму. При $p = 0,06$ по формуле (1.105) объемы выборок, гарантирующих наличие как минимум одного зараженного корнеплода на трех указанных уровнях достоверности суждения от P'_1 до P'_3 , будут 48, 74 и 112 корнеплодов.

§ 1.25. Нормальное распределение

Достаточно большую статистическую совокупность часто можно представить в виде эмпирической кривой распределения. Почти каждое эмпирическое распределение можно в свою очередь интерпретировать каким-либо теоретическим типом распределения, что, как правило, раскрывает новые особенности изучаемого явления и позволяет прогнозировать некоторые его свойства. Принципы построения статистических распределений и другие подробности о них можно найти в курсах теории вероятностей. Распределения разделяются на два основных типа: дискретные и непрерывные, в зависимости от характера изменения составляющих их случайных величин.

Нами, начиная с § 1.26, рассмотрены распределения дискретного типа: альтернативное, биномиальное, гипергеометрическое, Пуассона, негативное биномиальное, мультиномиальное, а затем, начиная с § 1.32, — распределения непрерывного типа: Максвелла, Рэлея, показательное, Парето и др. Предшествует тем и другим типам — нормальное распределение непрерывного типа, которое является основным, так как согласно теории вероятностей к нему в пределе стремятся все известные типы распределений. Нормальное распределение широко распространено, имеет наибольшее значение для теории и практики биометрических исследований, поэтому в первую очередь сравнивают полученное эмпирическое распределение с нормальным, теоретические частоты которого необходимо для этого вычислить. Нормальному распределению присущи следующие признаки: 1) симметричная колоколообразная двускатная кривая распределения частот; 2) соотношение средней арифметической и сигмы, равное $M = 3\sigma$, или $v = 33\%$. Чем ближе эмпирическое распределение к этим показателям, тем больше шансов на удовлетворительную аппроксимацию его кривой нормального распределения.

Итак, требуется установить, какие значения частот имел бы данный вариационный ряд, если распределение частот было бы нормальным. Теоретические частоты нормальной кривой можно вычислить по формуле:

$$f' = \frac{Nc}{\sigma \sqrt{2\pi}} \cdot e^{-0,5t^2}, \quad (1.106)$$

или, что одно и то же, по формуле:

$$f' = \frac{Nc}{\sigma} \cdot \frac{0,39894}{e^{0,5t^2}}, \quad (1.107)$$

или

$$f' = N \exp \{ \ln c + \ln 0,39894 - \ln \sigma - 0,5 t^2 \}, \quad (1.108)$$

где f' — теоретические частоты нормальной кривой; N — объем выборки; c — классовой интервал; σ — среднее квадратическое отклонение; e — основание натуральных логарифмов, равное 2,7182818 (см. табл. 22П); $t = (x - M)/\sigma$ — нормированное отклонение (о нем говорится также в § 1.23); M — средняя арифметическая; x — варианты (середины классов) ряда.

Формула (1.108) наиболее удобна при вычислении на небольших ЭВМ, так как дает возможность избежать переполнения памяти при больших показателях числа e .

Вычисление теоретических частот распределения 214 сортов ячменя по продолжительности вегетационного периода по данным из табл. 1.03, по формуле (1.106) показано в табл. 1.28, где произведены следующие действия.

1. Найдены разности вариантов и средней арифметической: $x - M$ (столбец 2).

2. Разности $x - M$ в столбце 3 нормируются, т. е. делятся на среднее квадратическое отклонение ($\sigma = 5,989$).

3. По вычисленным значениям t по табл. 5П находим значения выражения $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-0,5t^2}$ для формулы (1.106) или $\psi(t) = \frac{0,39894}{e^{0,5t^2}}$ для формулы (1.107) (столбец 4).

Таблица 1.28. Вычисление теоретических частот нормального распределения

Варианты, x	$x - M$	$t = \frac{x - M}{\sigma}$	$\psi(t)$ (табл. 5П)	Теоретические частоты, f'		Эмпири- ческие частоты,
				вычисленные	округленные	
60	-15,09	-2,5196	01667	2,98	3	2
65	-10,09	-1,6848	09728	17,4	17	30
70	-5,09	-0,8499	27798	49,7	50	34
75	-0,09	-0,0150	39889	71,3	72	62
80	4,91	0,8198	28504	50,9	51	74
85	9,91	1,6547	10139	18,1	18	8
90	14,91	2,4896	01797	3,2	3	4
$c = 5 \quad M = 75,09 \quad \sigma = 5,989$				213,58	214	214

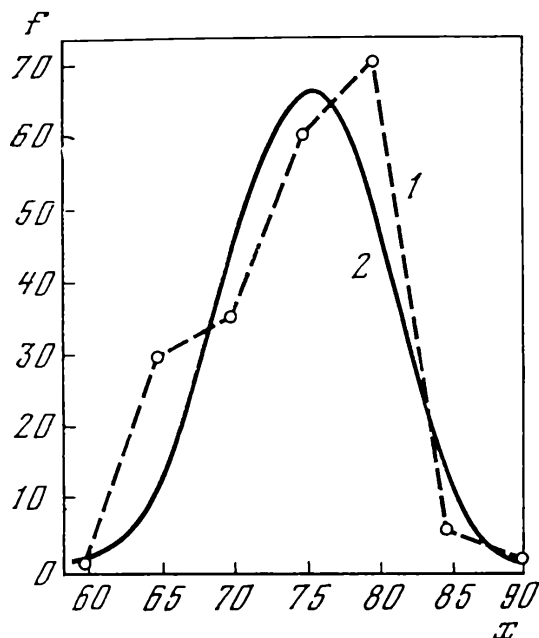


Рис. 3. Распределение сортов ячменя по продолжительности вегетационного периода

1 — эмпирические частоты;
2 — кривая нормального распределения;
f — частоты; x — варианты

На рис. 3 изображены эмпирическая и теоретическая кривые нормального распределения рассмотренной совокупности. Сравнение теоретических частот с эмпирическими с целью установить, отличается ли данное распределение от нормального, рекомендуется производить критерием Колмогорова, метод вычисления которого излагается в § 4.15. Величина этого критерия в данном случае составляет $K(\lambda) = 0,91$, что меньше величины первого критического уровня, равной 1,36 для доверительной вероятности 0,95. Следовательно, степень совпадения эмпирических и теоретических частот в нашем примере можно считать удовлетворительной. Теоретические частоты нормальной кривой в процентах можно приблизительно вычислить и для невзвешенного ряда, если принять в формуле (1.106) $N = 100\%$, $c = 10\%$ (что определяет число классов $k = 10$), начальную варианту $x_1 = 0$. Необходимость в таком расчете может возникнуть, например, при оценке степени нормальности распределения данных.

* * *

Одним из важных применений свойств нормального распределения служит возможность определения доли совокупности, находящейся в пределах заданного интервала изменения величины x , что покажем на следующем примере. Требуется выяснить, у какой части всех сортов ячменя урожай созревает на 70—80-й день после появления всходов, или, говоря иначе, какова вероятность встретить в культуре сорта ячменя с вегетационным периодом от 70 до 80 дней.

1. Для этого вычисляем нормированные отклонения меньшей и большей границ заданного интервала: $x_1 = 70$, $x_2 = 80$. Выше были

4. Вычисляем значение выражения:

$$\frac{Nc}{\sigma} = \frac{214,5}{5,989} = 178,66.$$

5. Множим числа столбца 4 на 178,66 (столбец 5).

6. Округляя числа столбца 5 до целых чисел, получаем теоретические частоты; их сумма должна быть близкой к объему выборки, а величина крайних вариантов, выраженных в нормированных отклонениях, должна быть не меньше 3. Если эти варианты меньше 3, то следует слева, справа или симметрично добавить несколько классов по краям распределения. Уравнение кривой нормального распределения в конкретной форме:

$$f' = 178,66\psi(t),$$

где $t = (x - 75,09)/5,989$.

вычислены средняя $M = 75,09$ дней и сигма $\sigma = 5,989$ дней. Отсюда нормированные отклонения равны:

$$t_1 = \frac{x_1 - M}{\sigma} = \frac{70 - 75,09}{5,989} = -0,8499;$$

$$t_2 = \frac{x_2 - M}{\sigma} = \frac{80 - 75,09}{5,989} = 0,8198.$$

2. По табл. 27П значений функции $\Phi(t)$, приведенной в приложении, находим по вычисленным t_1 и t_2 доли совокупности, находящиеся левее этих точек, но перед этим следует каждый раз преобразовывать отрицательные величины t и функции, если таковые имеются, в положительные. В нашем примере $t_1 = -0,8499$, поэтому функция для него равна:

$$\Phi(-0,8499) = 1 - [\Phi(0,8499)] = 1 - 0,8023 = 0,1977,$$

т. е. в случае отрицательного t найденное по таблице значение функции Φ вычитаем из единицы, остаток и будет искомым значением функции Φ для отрицательного t . Нормированное отклонение t_2 для второй, большей границы интервала — число положительное, поэтому находим сразу по таблице: $\Phi(0,8198) = 0,7939$.

3. По формуле $p = \Phi(t_2) - \Phi(t_1)$ находим искомую долю:

$$p = 0,7939 - 0,1977 = 0,5962, \text{ или } 59,62\%.$$

Таким образом, примерно у 60% всех известных сортов ячменя на созревание урожая требуется от 70 до 80 дней. Подобные сведения представляют интерес как при определении направления селекции новых сортов этой культуры, так и при географическом их размещении.

На основе свойств нормального распределения возможно решать также и обратные задачи: по заданной вероятности или объему совокупности в процентах или долях единицы найти количественную границу интервала, в котором эта доля будет находиться. При этом одна из границ интервала должна быть известна. Например, требуется определить, по какому количественному критерию следует отбирать 5% сортов ячменя, вегетационный период которых больше меди (средней арифметической).

Одна из границ искомого интервала известна — это значение меди: $M = 75,1$ день. Другую границу, численное значение которой больше меди, предстоит найти. Сначала решим эту задачу обычным способом при помощи таблицы функции нормального распределения. Сигма была вычислена раньше, она равна 5,989. Нормированное отклонение для средней арифметической, как известно, равно нулю: $t_1 = (75,1 - 75,1)/5,989 = 0$.

В табл. 17П находим, что этому значению соответствует вероятность 0,50, или 50%, всего объема совокупности. Далее в этом же столбце табл. 17П находим вероятность 0,55 ($0,50 + 0,05$, или 5%), которой соответствует $t_2 = 0,13$. Так как $t_2 = \frac{x - M}{\sigma} = \frac{x - 75,1}{5,989}$, то искомое значение второй границы интервала равно: $x = \sigma t_2 + M =$

Таблица 1.29. К вычислению доли совокупности по эмпирическим частотам

Вегетационный период ячменя, дни		Частота, f	Накопленная частота, f_{Σ}
Границы класса	Середина класса, x		
57,5—62,4	60	2	2
62,5—67,4	65	30	32
67,5—72,4	70	34	66
72,5—77,4	75	62	128
77,5—82,4	80	74	202
82,5—87,4	85	8	210
87,5—92,5	90	4	214
$k = 7$		$c = 5$	$N = 214$

$= 5,989 \cdot 0,13 + 75,1 = 75,9$ дней. Таким образом, для того чтобы отобрать 5% сортов с указанным признаком, надо выбрать из совокупности все те сорта, вегетационный период у которых равен 76 дням.

Решим эту же задачу вторым способом — непосредственно по эмпирическому ряду без помощи каких-либо таблиц нормального распределения, рассуждая следующим образом. 5% от 214 сортов составят: $214 \cdot 0,05 = 10,7 = 11$ сортов. По табл. 1.29 накопленных частот определяем, что значение медиан — одной из границ искомого интервала, находится в классе с границами от 72,5 до 77,4. Число сортов, приходящихся на одну единицу классического интервала: $(128 - 66)/5 = 52$ сорта. Определим прямолинейной интерполяцией, т. е. обычной пропорцией, сколько единиц классического интервала приходится на 11 сортов:

$$\begin{array}{l} 52 — 5 \text{ дней} \\ 11 — x_2 \end{array} \quad x_2 = \frac{11 \cdot 5}{52} = 1,03 \text{ дня.}$$

Таким образом, левая граница интервала округленно равна 75 дням, а правая: $75 + 1,06 = 76$ дням, что и было получено по первому способу решения этой задачи. Отобрав все сорта в пределах этого интервала, т. е. с вегетационным периодом 76 дней, получим 11 сортов, или 5% от их общего числа. В табл. 1.01 насчитывается 12 сортов с вегетационным периодом 76 дней, что можно считать хорошим совпадением результатов подсчета с фактическими данными. Только что изложенную задачу, решенную по эмпирическим частотам, можно выполнить и по следующим формулам, в зависимости от того, совпадают ли границы интервала искомой доли с границами или серединами класса.

1. Если границы интервала искомой доли совпадают с серединами классов распределения, например от $x_1 = 70$ до $x_2 = 80$ дней,

то:

$$p = \frac{f_{\Sigma 2} - f_{\Sigma 1} - 0,5(f_2 - f_1)}{N}, \quad (1.109)$$

где $f_{\Sigma 1}$, $f_{\Sigma 2}$ — накопленные частоты, соответствующие заданным x_1 и x_2 ; f_1 , f_2 — эмпирические частоты, соответствующие заданным x_1 и x_2 ; N — объем совокупности.

Подставляя известные данные из табл. 1.29, получим:

$$p = \frac{202 - 66 - 0,5(74 - 34)}{214} = 0,542,$$

что хотя и немного меньше доли, полученной через посредство таблиц функции $\Phi(t)$, однако, вероятно, более соответствует фактам, так как аппроксимация нормальной кривой, как это видно на рис. 3, могла бы быть лучшей.

2. Если границы интервала искомой доли совпадают с границами классов распределения, то:

$$p = \frac{f_{\Sigma 2} - f_{\Sigma 1}}{N}, \quad (1.110)$$

где обозначения те же.

Например, для интервала от $x_1 = 72,5$ до $x_2 = 82,4$ дней доля совокупности, или вероятность, равна:

$$p = \frac{202 - 128}{214} = 0,346, \text{ или } 34,6\% \text{ сортов.}$$

3. Если границы интервала искомой доли не совпадают ни с границами классовых интервалов, ни с серединами классов, то вычисления доли (p) ведем следующим образом. Находим частоты f'_1 и f'_2 , соответствующие, например, таким заданным границам интервала доли: $x_1 = 64$, $x_2 = 73$, для чего делим частоту, соответствующую меньшей границе интервала доли (x_1), на классовой интервал. Получаем число частот, приходящееся на единицу классовой интервала: $30/5 = 6$. Умножаем его на разницу x_1 с меньшей границей класса ($64 - 62,5$), т. е. $6 \cdot 1,5 = 9$, и вычитаем полученное число из частоты f'_1 класса, в котором находится x_1 : $30 - 9 = 21$, что можно представить в виде формулы:

$$f'_1 = f_1 \left(1 - \frac{x_1 - G_1}{c} \right) = 30 \left(1 - \frac{64 - 62,5}{5} \right) = 21, \quad (1.111)$$

где f'_1 — частота, точно соответствующая x_1 — меньшей границе интервала искомой доли; f_1 — частота класса, в котором находится x_1 ; G_1 — меньшая граница класса, в котором находится x_1 ; c — классовой интервал.

Частота для x_2 — большей границы интервала доли равна:

$$f'_2 = \frac{f_2}{c} (x_2 - G_2), \quad (1.112)$$

где G_2 — меньшая граница класса, в котором находится значение

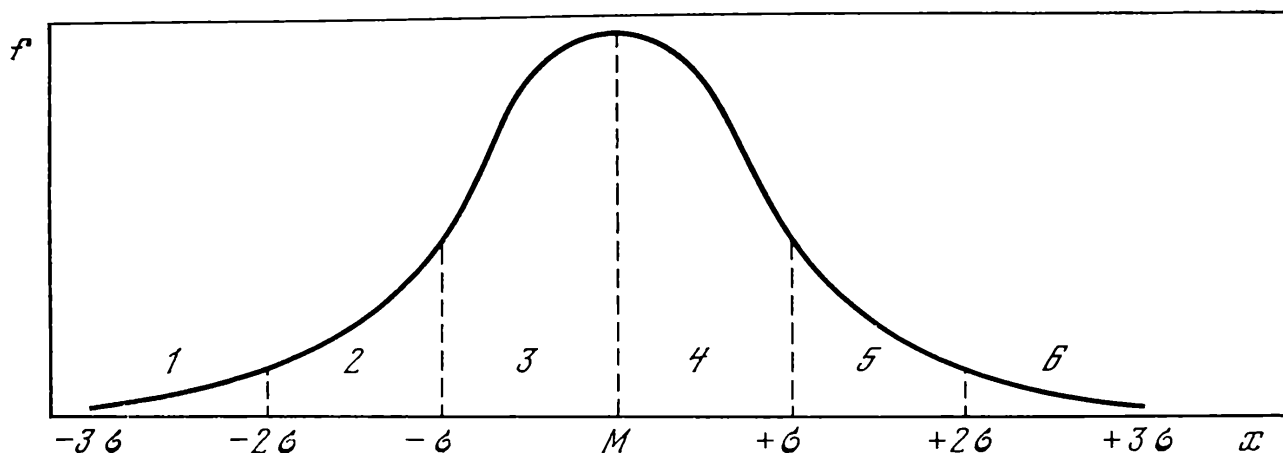


Рис. 4. Нормальная кривая

На оси абсцисс — баллы шкалы по среднему квадратичному отклонению; M — средняя арифметическая; f — частоты; x — варианты

x_2 , а f_2 — частота этого класса, т. е.

$$f'_2 = \frac{62}{5} (73 - 72,5) = 6,2.$$

Из таблицы накопленных частот данного ряда находим $f_{\Sigma 1} = 32$; $f_{\Sigma 2} = 128$, после чего вычисляем искомую долю по формуле:

$$p = \frac{f_{\Sigma 2} - f_{\Sigma 1} + f'_1 - f_2 + f'_2}{N}, \quad (1.113)$$

или:

$$p = \frac{128 - 32 + 21 - 62 + 6,2}{214} = 0,286,$$

т. е. среди сортов ячменя насчитывается 28,6% созревающих в период от 64 до 74 дней после посева.

При количественном определении понятий «типичное» и «норма» важную роль играет среднее квадратическое отклонение. На расстоянии $M + \sigma$ и $M - \sigma$ от медики на графике нормальной кривой расположены абсциссы ее двух точек перегиба (рис. 4). Точки перегиба кривых, как правило, указывают на тот момент процесса или явления, когда в нем происходят качественные изменения, что важно для построения теоретически обоснованных шкал. В нормальном распределении в точках перегиба может произойти лишь переход от типичных величин вариантов совокупности к нетипичным, хотя и принадлежащих еще к данной совокупности. В этом, собственно, и заключается смысл среднего квадратического отклонения, которое измеряет среднюю величину отклонений вариантов от медики, наиболее вероятной или типичной величины.

Приведем величины площади под нормальной кривой, которые соответствуют определенным значениям сигмы, одновременно по обе стороны от медики. В интервале нормы, между абсциссами, от $M - \sigma$ до $M + \sigma$ находится 68,27% всей площади нормального распределения, т. е. вариант, или дат совокупности; между $M - 2\sigma$

Таблица 1.29а. Доли вариант, заключенных в различных сигмальных интервалах

t	Границы интервалов		Доли вариант	
		до	при нормальном распределении	минимум, при любом распределении
1	$M - \sigma$	$M + \sigma$	0,683	0
1,5	$M - 1,5\sigma$	$M + 1,5\sigma$	0,866	0,555
2	$M - 2\sigma$	$M + 2\sigma$	0,955	0,750
2,5	$M - 2,5\sigma$	$M + 2,5\sigma$	0,988	0,840
3	$M - 3\sigma$	$M + 3\sigma$	0,997	0,889
3,5	$M - 3,5\sigma$	$M + 3,5\sigma$	0,999	0,918

и $M + 2\sigma$ заключается 95,45% дат от всего объема и в интервале от $M - 3\sigma$ до $M + 3\sigma$ лежит 99,73% от всего объема нормально распределенной совокупности.

Если тип распределения данных неизвестен или известно, что он не относится к нормальному, то минимальную долю вариант, входящих на интервал $M \mp t$, можно определить по формуле:

$$p_{\min} > 1 - \frac{1}{t^2}, \quad (1.113a)$$

где p_{\min} — минимальная доля вариант, заключающаяся в интервале от $-t$ до $+t$; $\pm t$ — нормированное отклонение, взятое со знаком $+$ и $-$, т. е. находящееся на одинаковом расстоянии от средней арифметической слева и справа по оси абсцисс. При определении доли вариант в сигмальном интервале ($\mp \sigma$), когда $\pm t = 1$, $p_{\min} > 1 - 1/1 > 0$. Таким образом, не исключено, что в интервале нормы при некотором типе распределения может оказаться любое число вариант, о котором известно лишь, что оно больше нуля. При $t < 1$ в правой части приведенного неравенства получается отрицательное число. Следовательно, в интервале меньше $\pm \sigma$ по обе стороны от средней при некоторых типах распределений не гарантируется наличие даже одной варианты.

Доля вариант при нормальном и любом другом распределении данных, для некоторых значений t приведена в табл. 1.29а.

§ 1.26. Альтернативное распределение

Альтернативное (или дихотомическое) распределение применяется в тех случаях, когда варианты могут быть разделены по некоторому признаку только на два класса. Например, совокупность корнеплодов моркови может быть разделена на пораженные и непораженные фомозом; совокупность цветков на однодомном растении может быть разделена на мужские и женские, и т. д. Таким образом, при альтернативном варьировании дат совокупность делится на две части: в одной из них наблюдаемый признак, например поражение болезнью, присутствует, а в другой части наблюдаемый признак отсутствует.

Выборки, состоящие из двух групп альтернативных вариантов, нередко встречаются в практике биологических, в частности, генетических исследований. Например, в табл. 1.04 в одной выборке, состоящей из 1050 обследованных растений ячменя, 66 особей оказались мутантами, что составляет 6,3% от всего числа растений, или 0,063 в долях единицы. Доли особей, или вариант, обладающих изучаемым признаком или не обладающих им, определяются по формулам:

$$p = n_p/N, \quad (1.114)$$

$$q = n_q/N, \quad (1.115)$$

где p — доля особей с изучаемым признаком, выраженная как часть от единицы, она же и средняя арифметическая; q — доля особей без этого признака; n_p — число особей, имеющих изучаемый признак; n_q — число особей, не имеющих изучаемого признака; N — объем всей выборки ($N = n_p + n_q$).

Когда доли выражаются в частях от единицы, всегда имеет место следующее равенство:

$$p + q = 1, \quad (1.116)$$

которое служит для проверки вычислений.

Доли особей с изучаемым признаком или без него могут быть определены также в процентах по формулам:

$$p = \frac{n_p}{N} \cdot 100\%, \quad (1.117)$$

$$q = \frac{n_q}{N} \cdot 100\%. \quad (1.118)$$

К формулам (1.117), (1.118) также существует равенство:

$$p + q = 100\%, \quad (1.119)$$

где обозначения аналогичны тем, которые даны к формулам (1.114)—(1.116). Для данных табл. 1.04 доли мутантов и немутантов в частях единицы определяются по формулам (1.114)—(1.116):

$$p = \frac{66}{1050} = 0,063;$$

$$q = \frac{984}{1050} = 0,937;$$

$$p + q = 0,063 + 0,937 = 1,$$

или в процентах по формулам (1.117)—(1.119):

$$p = \frac{66}{1050} \cdot 100 = 6,3\%,$$

$$q = \frac{984}{1050} \cdot 100 = 93,7\%, \quad p + q = 6,3 + 93,7 = 100\%.$$

Для определения достоверности доли при альтернативном разделении вариант необходимо определить ее ошибку по формуле:

$$m_M = \sqrt{\frac{pq}{N}}, \quad (1.120)$$

где m_M — ошибка доли; p — доля вариант, имеющих наблюдаемый признак, или средняя арифметическая этого признака; q — доля вариант, не имеющих наблюдаемого признака; N — объем выборки.

Для данных табл. 1.04:

$$m_M = \sqrt{\frac{6,3 \cdot 93,7}{1050}} = \pm 0,75 \%$$

Ошибка получается в долях единицы или в процентах в зависимости от размерности p и q . Границы достоверности при альтернативном разделении вариант можно определить, прибавляя к той или другой доле выражение $\pm t m_M$, где t — критерий Стьюдента (по табл. ЗП) с числом степеней свободы $\nu = N - 1$. Для рассматриваемого случая, при $P'_1 = 0,95$, $N = 1050$, $\nu = 1050 - 1 = 1049$, $t = 1,96$ (из табл. ЗП) $p = 6,3\% \pm 1,96 \cdot 0,75 = 6,3 \pm \pm 1,5\%$, т. е. процент мутантов в совокупности особей ячменя при воздействии на них некоторых мутагенных факторов, или средняя арифметическая мутантов, может колебаться от 4,8 до 7,8%. Ошибка альтернативной доли равна той же величине, т. е. $m_q = m_M$, в данном случае: $q = 93,7\% \pm 1,5\%$.

При $p = 0$ ошибку доли вычисляют при помощи искусственного приема по формулам:

$$p = \frac{(n_p + 1) 100}{N + 2}, \quad (1.121)$$

$$m_m = \sqrt{\frac{pq}{N + 3}},$$

где обозначения те же, что и к формулам (1.114), (1.120).

Допустим, что в другой выборке из 997 растений ячменя не оказалось ни одного мутанта, т. е. $n_p = 0$. Тогда:

$$p = \frac{(0 + 1) 100}{997 + 2} = 0,1\%, \quad m_m = \sqrt{\frac{0,1 \cdot 99,9}{997 + 3}} = 0,1\%,$$

т. е. в генеральной совокупности возможно все же встретить небольшой процент мутантов, несмотря на их отсутствие в данной выборке. Формулы (1.121), конечно, применимы также и к альтернативной доле — q .

Среднее квадратическое отклонение, или сигма двухклассового разделения вариант, определяется для количественного выражения меры альтернативной изменчивости в долях единицы или процентах по формуле:

$$\sigma = \sqrt{pq}, \quad (1.122)$$

где σ — среднее квадратическое отклонение; p — доля вариант, обладающих наблюдаемым признаком, в долях единицы или процентах; q — доля вариант, не обладающих наблюдаемым признаком.

Для данных из табл. 1.04 сигма в долях единицы по формуле (1.122):

$$\sigma = \sqrt{0,063 \cdot 0,937} = 0,243.$$

или в процентах:

$$\sigma = \sqrt{6,3 \cdot 93,7} = 24,296 \%$$

Ошибку сигмы можно определить по формуле (1.32). Величина сигмы при альтернативном варьировании не может превышать 50%, или 0,5 в долях единицы.

Для альтернативного варьирования дат коэффициент вариации можно вычислить по формуле:

$$v = 2\sigma \%, \quad (1.123)$$

где v — коэффициент вариации; σ — среднее квадратическое отклонение, в процентах.

При среднем квадратическом отклонении, равном 24,296:

$$v = 2 \cdot 24,296 = 48,6 \%$$

Ошибку коэффициента вариации с достаточной точностью можно определить по формуле (1.47).

С достаточным приближением показатель точности опыта при альтернативном разделении дат определяется по формулам (1.50) и (1.51). Для рассмотренного примера показатель точности опыта по формуле (1.51):

$$P = \frac{48,6}{\sqrt{1050}} = 1,5 \%$$

Такое значение является высоким показателем точности определения доли, или средней арифметической, в данном опыте. Точность опыта считается неудовлетворительной, если показатель P больше 5%.

§ 1.27. Биномиальное распределение

Теоретические частоты распределения вычисляют главным образом для того, чтобы путем сравнения с эмпирическими частотами выяснить меру их совпадения. При совпадении опытных и теоретических частот на эмпирическую выборку распространяют закономерности, присущие данному теоретическому типу распределения.

Для применения биномиального распределения исходные данные должны быть дискретного типа и отвечать следующим условиям.

1. Объем исходной совокупности, из которой делается выборка, должен быть настолько велик, чтобы не могла измениться вероятность определения данного признака. Если объем исходной (генеральной) совокупности недостаточно велик, то отобранные единицы учета должны быть возвращены в нее для сохранения той же вероятности получения выборки с данным признаком. Если совокупность не велика и отобранные единицы не могут быть в нее возвращены, то в этом случае следует применить гипергеометрическое распределение (§ 1.28).

2. Величина дисперсии выборки должна быть меньше ее средней арифметической. Выборка при биномиальном распределении обычно

Таблица 1.30. Вычисление частот биномиального распределения по формуле (1.107)

Число пора- женных кор- неплодов в пробе, x	Число проб, f	xf	x^2f	$t = \frac{x - M}{\sigma}$	$\psi(t)$	$f' = \frac{N_n}{\sigma} \psi(t)$
9	4	36	324	1,802	0,07895	5
8	20	160	1280	1,255	0,18037	11
7	18	126	882	0,708	0,31006	19
6	14	84	504	0,161	0,393870	24
5	27	135	675	0,386	0,371150	23
4	18	72	288	0,932	0,258880	16
3	4	12	36	1,479	0,133440	8
2	7	14	28	2,026	0,05082	3
1	0	0	0	2,572	0,014603	1
0	0	0	0	3,119	0,0030698	0
112		639	4017			110

образуется, когда берут N_n проб одинакового объема, равного n . В каждой из проб подсчитывают число вариантов, обладающих изучаемым признаком. Например, было взято $N_n = 112$ проб корнеплодов из совокупности сортов моркови с целью изучения их устойчивости к заболеванию фомозом. Корнеплоды заражались искусственно с механическим их повреждением. В каждой из 112 проб было одинаковое число корнеплодов: $n = 9$. В каждой девятке подсчитывалось число пораженных фомозом корнеплодов (табл. 1.30). Соответствие эмпирического распределения биномиальному можно определить двумя способами в зависимости от величины пробы n . Если n велико, например больше 10, теоретические частоты биномиального распределения можно получить с достаточной точностью при помощи формулы частот нормального распределения. Если $n < 10$, лучше воспользоваться формулой разложения бинома, отдельные члены которого представляют собой теоретические частоты биномиального распределения. Оба способа рассмотрим на данных из столбцов 1 и 2 табл. 1.30.

Теоретические частоты биномиального распределения по первому способу, по схеме нормального распределения, вычисляются по формуле (1.107):

$$f' = \frac{N_n \psi(t)}{\sigma},$$

где сигма может быть найдена, например, по формуле (1.38):

$$\sigma = \sqrt{\frac{\sum fx^2 - (\sum fx)^2/N_n}{N_n - 1}}.$$

Обозначения в этих формулах следующие: f' — теоретические частоты биномиального распределения; N_n — число проб; $\psi(t)$ —

функция от t , отыскивается по табл. 5П; t — нормированное отклонение, $t = \frac{x - M}{\sigma}$; σ — среднее квадратическое отклонение; x — варианты распределения; M — средняя арифметическая; f — эмпирические частоты распределения; c — классовый интервал.

Теоретические частоты по первому способу, приведенному в табл. 1.30, вычисляем в следующем порядке.

1. Перемножаем варианты из столбца 1 на соответствующие им частоты из столбца 2, результаты пишем в столбце 3.

2. Числа столбца 3 умножим на числа столбца 1, результаты записываем в столбце 4.

3. Суммируем числа столбцов 2—4, получаем итоги: $N_n = 112$, $\Sigma xf = 639$, $\Sigma x^2 f = 4017$.

4. Вычисляем среднюю арифметическую (1.13):

$$M = 639/112 = 5,705 \text{ корнеплода.}$$

5. Вычисляем среднее квадратическое отклонение (1.38):

$$\sigma = \sqrt{\frac{4017 - 639^2/112}{112 - 1}} = 1,829.$$

Дисперсия $\sigma^2 = 3,34$ меньше, чем средняя, поэтому в данном случае можно применить биномиальное распределение.

6. Разности средней арифметической с вариантами x делим на $\sigma = 1,829$, получаем значения нормированных отклонений в столбце 5.

7. По табл. 5П находим значения функции $\psi(t)$ и выписываем их в столбец 6.

8. Находим значение выражения:

$$\frac{N_n c}{\sigma} = \frac{112 \cdot 1}{1,829} = 61,236.$$

9. Перемножаем числа из столбца 6 на 61,236 и, округляя до целых чисел, получаем в столбце 7 теоретические частоты кривой биномиального распределения (рис. 5), уравнение которой:

$$f' = 61,236 \psi(t), \text{ где } t = \frac{x - 5,7}{1,8}$$

Для вычисления теоретических частот биномиального распределения по второму способу используем формулу разложения бинома:

$$f' = N_n (p + q)^n = N_n \left[p^0 q^n + n p^1 q^{n-1} + \frac{n(n-1)}{1 \cdot 2} p^2 q^{n-2} + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} p^3 q^{n-3} + \dots + p^n q^0 \right], \quad (1.124)$$

где p — доля вариантов, обладающих данным признаком; q — доля вариантов, не обладающих данным признаком; n — число вариантов в пробе; N_n — число проб.

В рассматриваемом примере (табл. 1.30) число пораженных корнеплодов $n_p = 639$ (итог столбца 3), а общее число корнеплодов — $N = N_n \cdot n = 112 \cdot 9 = 1008$. Следовательно, $n_q = 1008 - 639 =$

= 369. В долях единицы от общего числа корнеплодов: $p = 0,63$; $q = 0,37$. Размер пробы $n = 9$. Общее число проб $N_n = 112$. Вычисление теоретических частот по формуле (1.124): $f' = (0,63 + 0,37)^9 \cdot 112$, ведется в следующем порядке (табл. 1.31).

1. Из табл. 6П выписываем в столбце 1 коэффициенты разложения бинома при $n = 9$ (размер пробы).

2. Возводим в степени от 0 до 9 долю $p = 0,63$ (столбце 2).

3. Возводим в степени от 0 до 9 долю $q = 0,37$ (столбце 3).

4. Перемножаем по строкам числа столбцов 2 и 3, произведения записываем в столбец 4.

5. В столбце 5 получаем произведения чисел столбцов 1 на 4.

6. В столбце 6 записываем произведения числа проб $N_n = 112$, на числа из столбца 5, произведения округляем до целых чисел, которые и будут теоретическими частотами биномиального распределения.

Если $p = q = 0,5$, то кривая биномиального распределения будет симметричной.

Вычисление частостей биномиального распределения наиболее легко производится по рекуррентной (т. е. когда при вычислении

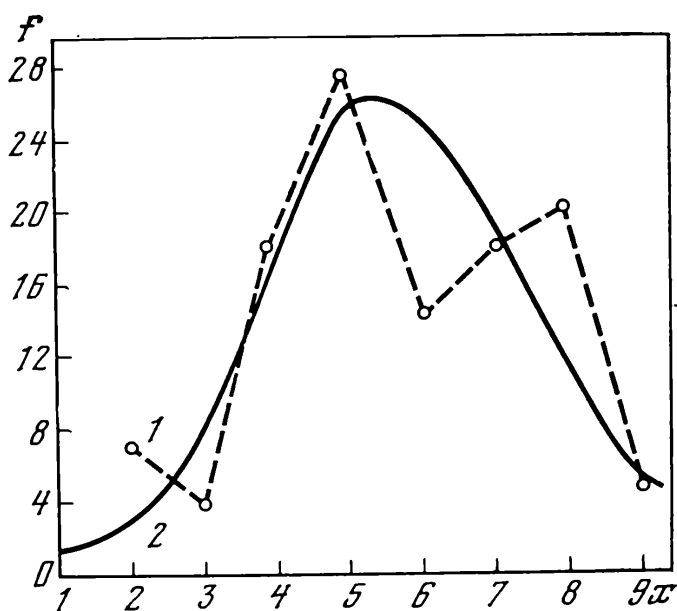


Рис. 5. Эмпирические (1) и теоретические (2) частоты (f) биномиального распределения поражаемости фомозом корнеплодов моркови

-- варианты

Таблица 1.31. Вычисление частот биномиального распределения по формуле (1.124)

Число из табл. 6П при $n = 9$	Степени		pq	Произведение чисел столбцов 1-4	Произведение $N_n = 112$ на числа столбца 5 f'	Частоты, %
	$p = 0,63$	$q = 0,37$				
1	$p^9 = 0,0156$	$q^0 = 1,00$	0,0156	0,0156	2	1,8
9	0,0248	0,37	0,0092	0,0826	9	8,0
36	0,0394	0,1369	0,0054	0,1942	22	19,6
84	0,0625	0,0507	0,0032	0,2662	30	26,8
126	0,0992	0,0187	0,0019	0,2337	26	23,2
126	0,1575	0,0069	0,0011	0,1369	16	14,3
84	0,2501	0,0026	0,0007	0,0546	6	5,4
36	0,3969	0,001	0,0004	0,0143	2	0,9
9	0,63	0,0004	0,0003	0,0023	0	
1	$p^0 = 1$	$q^9 = 0,0001$	0,0001	0,0001	0	
					112	100,0

последующего используется предыдущее число) формуле:

$$f'_{i+1} = \frac{n-i}{i+1} \cdot \frac{p}{q} f'_i, \quad (1.125)$$

где f'_{i+1} — последующая частота; n — объем пробы; p, q — доли присутствия и отсутствия изучаемого признака в совокупности; f'_1 — предыдущая частота; i — номер члена ряда частот.

Заметим, что рекуррентные формулы требуют достаточной точности вычислений, так как все последующие результаты расчетов основаны на предыдущих и даже небольшая ошибка, допущенная вначале, быстро увеличивается в конце ряда.

В нашем примере: $n = 9, p = 0,63, q = 0,37$. Начальный член (f'_0) ряда частот находим по формуле:

$$f'_0 = q^n, \quad (1.126)$$

обычно через любые (десятичные или натуральные) логарифмы: $\ln f'_0 = n \ln q$, или $f'_0 = \exp(n \ln q)$, т. е. для нашего примера: $f'_0 = \exp(9 \ln 0,37) = \exp[9(-0,99425)] = \exp(-8,94825) =$

$= 0,00012996$. Последующие члены вычисляем по формуле (1.125), например: $f'_1 = \frac{9}{1} \cdot \frac{0,63}{0,37} \cdot 0,00012996 = 0,0019915$ и т. д.

(см. табл. 1.32).

В табл. 1.32 величинам изучаемого признака — от 1 до 9 больных корнеплодов ($n - i$, в столбце 1) соответствуют частоты биномиального распределения (в %) от 1,56 до 0,199 (столбец 6), которые могут быть нанесены на график. Оценка посредством критерия Колмогорова (см. § 4.15) меры совпадения частот вычисленных и эмпирических показала, что лучшее совпадение в рассматриваемом случае дает применение уравнения нормальной кривой [$K(\lambda) = 0,76$]; расчеты по рекуррентной формуле [$K(\lambda) = 1,22$] и по

Таблица 1.32. Вычисление частот биномиального распределения по рекуррентной формуле (1.125)

$n-i$	$i+1$	$\frac{n-i}{i+1}$	$(3) \cdot \frac{p}{q}$	$f'_{i+1} = (4) \cdot f'_i$	$f' \%$		$f' = (5) \cdot N_n$
9	1	9	15,3243	0,0019915	0,199	0	0,015
8	2	4	6,8108	0,013564	1,36	1	0,223
7	3	2,333	3,9724	0,053882	5,39	2	1,52
6	4	1,5	2,5541	0,13762	13,8	3	6,04
5	5	1	1,7027	0,23433	23,4	4	15,4
4	6	0,6667	1,1352	0,26601	26,6	5	26,2
3	7	0,4286	0,7298	0,19423	19,4	6	29,8
2	8	0,25	0,4257	0,082684	8,27	7	21,7
1	9	0,1111	0,1892	0,015644	1,56	8	9,26
						9	1,75

99,98 $N_n = 112 \ 111,91$

$p/q = 0,63/0,37 = 1,7027$;

$f'_0 = 0,37^9 =$
 $= 0,00012996$

формуле разложения бинома $[K(\lambda) = 1,13]$ дают примерно одинаковые результаты.

Теоретические частоты биномиального распределения показывают, какова вероятность появления проб с данным числом в ней пораженных корнеплодов. В табл. 1.31, столбце 7 указано, сколько проб в процентах от общего их количества может заключать в себе то или иное число пораженных корнеплодов. Так, со всеми 9 пораженными корнеплодами может встретиться 1,8% всех проб; появление проб с 1 или 0 пораженных корнеплодов также маловероятно. Оценка совпадения теоретических частот распределения с эмпирическими производится в основном по критерию хи-квадрат (см. § 4.14).

Среднее квадратическое отклонение биномиального распределения в именованных числах (например, в числе особей):

$$\sigma = \sqrt{npq}, \quad (1.127)$$

где σ — среднее квадратическое отклонение; n — объем пробы (см. в начале этого параграфа); p — доля наблюдаемого признака в совокупности; q — доля особей без наблюдаемого признака.

Средняя арифметическая биномиального распределения, выраженная в именованных числах равна:

$$M = np, \quad (1.128)$$

где M — средняя арифметическая для пробы; n — объем пробы; p — доля наблюдаемого признака.

При изучении повреждаемости фомозом корнеплодов моркови брали $N_n = 112$ проб по 9 корнеплодов ($n = 9$) в каждой; доля поврежденных корнеплодов оказалась $p = 0,63$, что составляет в среднем по формуле (1.128):

$$M = 9 \cdot 0,63 = 5,67 \text{ корнеплода в пробе.}$$

Среднее квадратическое отклонение для этого примера равно (1.127): $\sigma = \sqrt{9 \cdot 0,63 \cdot 0,37} = 1,45$ корнеплода в среднем для одной пробы.

Указанные параметры могут быть выражены также и в относительных числах, по примеру того, как это сделано в § 1.26. Ошибка средней арифметической биномиального распределения приближенно определяется по формуле (1.120). Показатели асимметрии (A_s) и эксцесса (E) в рассматриваемом распределении:

$$A_s = \frac{q - p}{\sqrt{Npq}}, \quad (1.129)$$

$$E = \frac{(pq)^{-1} - 6}{N}, \quad (1.130)$$

где остальные обозначения те же, что и к формуле (1.124)

В нашем примере:

$$A_s = \frac{0,37 - 0,63}{\sqrt{1008 \cdot 0,63 \cdot 0,37}} = -0,017;$$

$$E = \frac{(0,63 \cdot 0,37)^{-1} - 6}{1008} = -0,0017$$

Чем ближе к нулю величины A_s и E , а p — к 0,5, тем симметричнее кривая данного распределения частот. При $p < 0,5$ величина показателя асимметрии положительна, а при $p > 0,5$ — отрицательна. Центральные моменты биномиального распределения равны: $\mu_1 = 0$, $\mu_2 = Npq$, $\mu_3 = Npq(q - p)$, $\mu_4 = Npq[3pq(N - 2) + 1]$.

$$(1.131)$$

Коэффициент вариации:

$$v = \sqrt{\frac{q}{Np}}. \quad (1.132)$$

Кроме перечисленных способов частоты биномиального распределения вычисляют также и при помощи таблиц. Рассмотрим это на примере распределения частот неразвитых семян яблони домашней, которых насчитали 148 из $N = 307$ (по неопубликованным данным С. Ф. Демидовой). Следовательно: $p = 148/307 = 0,48208$; $q = 159/307 = 0,51792$, где последнее число обозначает долю нормальных, выполненных семян. Выберем число классов, например $k = 10$. По табл. 5.1 в книге: Л. Н. Большева и Н. В. Смирнова, «Таблицы математической статистики» (1968) для этого числа классов и при $p = 0,5$ (ближайшее число в таблице) находим теоретические частоты (в % с округлением): 0,1; 1,0; 4,4; 11,7; 20,5; 24,6; 20,5; 11,7; 4,4; 1,0; 0,1 соответственно для i от 1 до 10. Если будем брать из этой совокупности пробы по 10 семян в каждой, то со всеми выполненными семенами найдем 0,1% от числа всех проб, с одним неразвитым семенем — 1,0%, с двумя неразвитыми семенами — 4,4% и т. д. Как видим, по этому способу можно получить ряд теоретического распределения, зная только долю признака. Среднее квадратическое отклонение в данном примере: $\sigma = \sqrt{307 \cdot 0,5 \cdot 0,5} = 8,76$ семян.

Приведем еще один пример работы с этими же таблицами. При анализе партии семян лещины из 79 семян оказалось 8 невсхожих, что составляет $p = \frac{100 \cdot 8}{79} = 10,13\%$. При числе классов $k = 5$ частоты этого распределения (в % с округлением) для вариантов от 0 до 5 соответственно равны: 59; 33; 7; 1; 0,05; 0,001%.

§ 1.28. Гипергеометрическое распределение

Некоторые эмпирические дискретные распределения биологических признаков в зависимости от принципа организации выборки можно моделировать гипергеометрическим распределением вместо биномиального. Дело в том, что последний тип распределения требует соблюдения принципа случайного отбора проб из генеральной совокупности, для чего следует располагать очень большим объемом исходного материала или же каждый раз возвращать пробу обратно в исходную совокупность и тщательно перемешивать объекты учета для того, чтобы вероятность доли содержания учитываемого признака в совокупности от пробы к пробе оставалась одной и той же.

Гипергеометрическое распределение предоставляет возможности учитывать объем совокупности, а также не возвращать пробы обратно в совокупность и не перемешивать ее, что на практике более удобно.

Приведем два способа расчета частостей этого распределения. Отдельно взятая частость гипергеометрического распределения для любого значения x вычисляется по формуле:

$$f' = \frac{C_{n_p}^x C_{n_q}^{n-x}}{C_N^n}, \quad (1.133)$$

где f' — частость гипергеометрического распределения в долях единицы; $C_{n_p}^x = \frac{n_p (n_p - 1) (n_p - 2) \dots (n_p - x + 1)}{1 \cdot 2 \cdot 3 \dots x}$; x — варианта ряда; n_p — число вариантов, обладающих учитываемым признаком; $C_{n_q}^{n-x} = \frac{n_q (n_q - 1) (n_q - 2) \dots (n_q - n + x + 1)}{1 \cdot 2 \cdot 3 \dots (n - x)}$; n — объем пробы; n_q — число вариант в совокупности, не обладающих учитываемым признаком; N — объем совокупности; $C_N^n = \frac{N (N - 1) (N - 2) \dots (N - n + 1)}{1 \cdot 2 \cdot 3 \dots n}$.

В распределении числа зараженных корнеплодов моркови, приведенном в табл. 1.30: $n_p = 639$, $n = 9$, $n_q = 369$, $N = 1008$ корнеплодов. Рассчитаем для $x = 3$ значение его частости, т. е. определим вероятность найти 3 зараженных в пробе из 9 корнеплодов. По формуле (1.133)

$$f'_3 = \frac{\frac{639 \cdot 638 \cdot 637}{1 \cdot 2 \cdot 3} \cdot \frac{369 \cdot 368 \cdot 367 \cdot 366 \cdot 365 \cdot 364}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}}{\frac{1008 \cdot 1007 \cdot 1006 \cdot 1005 \cdot 1004 \cdot 1003 \cdot 1002 \cdot 1001 \cdot 1000}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9}} = 0,051,$$

или 5,1%. Вычисления проводим через посредство логарифмов, предварительно сократив дробь. Сумма десятичных логарифмов числителя в данном примере равна 26,5018, а знаменателя — 27,7937,

Таблица 1.33. Эмпирические частоты и частости (f , $f\%$ по данным табл. 1.30) и частости гипергеометрического распределения (f' , $f'\%$)

	f	$f\%$	f'	$f'\%$
0	0	0	0,00011091	0,0111
1	0	0	0,0017669	0,177
2	7	6,2	0,012456	1,25
3	4	3,6	0,051002	5,10
4	18	16,1	0,13367	13,4
5	27	24,1	0,23255	23,2
6	14	12,5	0,26855	26,9
7	18	16,1	0,19179	19,2
8	20	17,8	0,082345	8,23
9	4	3,6	0,015645	1,56
	112	100	0,989886	99,0281

их разница, т. е. результат деления числителя на знаменатель, равна логарифму — 1,2919 или в искусственной форме логарифма $\bar{2},7081$, что соответствует числу 0,051. Вычислять по формуле (1.133) рекомендуется не менее чем с 5 знаками мантиссы логарифмов, особенно в тех случаях, когда объем выборки больше 1000.

Второй способ расчета частостей рассматриваемого типа распределения следует применять тогда, когда желательно получить все теоретические частоты для аппроксимации эмпирического ряда.

Проведем этот расчет для тех же данных. В табл. 1.33 приведены из табл. 1.30 варианты (столбец 1), их эмпирические частоты (столбец 2), частоты (столбец 3), а также теоретические частоты гипергеометрического распределения (столбцы 4 и 5).

1. Определим исходное значение частоты для $x = 0$. Это значение следует вычислять с 6—7 знаками после запятой, так как оно послужит начальной величиной в цепочке дальнейших расчетов.

$$f'_0 = \frac{C_{n_q}^n}{C_N^n} = \frac{n_q(n_q - 1)(n_q - 2) \dots (n_q - n + 1)}{N(N - 1)(N - 2) \dots (N - n + 1)} \quad (1.134)$$

В формуле (1.134) обозначения те же, что и в формуле (1.133); f'_0 — частость начального, нулевого члена распределения.

Подставляя известные величины в формулу (1.134), получим через натуральные логарифмы частость нулевого члена:

$$\begin{aligned} \ln f'_0 &= \ln \frac{369 \cdot 368 \cdot 367 \cdot 366 \cdot 365 \cdot 364 \cdot 363 \cdot 362 \cdot 361}{1008 \cdot 1007 \cdot 1006 \cdot 1005 \cdot 1004 \cdot 1003 \cdot 1002 \cdot 1001 \cdot 1000} = \\ &= 11,65231932 - 20,75916426 = -9,10684494, \end{aligned}$$

что соответствует $f'_0 = 0,00011091$, или 0,0111%.

Через десятичные логарифмы: $\lg f'_0 = 23,0605 - 27,0149 = -3,9544$, или в искусственной форме $\lg f'_0 = \bar{4},0456$, что соответствует $f'_0 = 0,0001111$, или 0,0111%.

Вычисление следующих членов ряда частостей ведется по рекуррентной формуле:

$$f'_i = \frac{(n - x_{i-1})(n_p - x_{i-1})f'_{i-1}}{(x_{i-1} + 1)(N - n_p - n + x_{i-1} + 1)}, \quad (1.135)$$

где f'_i — частость для x_i ; x_{i-1} — предыдущая варианта; n_p — число вариантов с учитываемым признаком; N — объем выборки; n — объем пробы (число классов распределения); f'_{i-1} — частость предыдущей варианты.

Последовательно вычисляя по формуле (1.135), получим все остальные частоты.

$$f'_1 = \frac{(9 - 0)(639 - 0)}{(0 + 1)(1008 - 639 - 9 + 0 + 1)} \cdot 0,00011091 = 0,0017669.$$

Заметим, что $N - n_p - n + 1 = 1008 - 639 - 9 + 1 = 361$,

в данном примере величина постоянная.

$$f'_2 = \frac{(9-1)(639-1)}{(1+1)(361+1)} \cdot 0,0017669 = 0,012456;$$

$$f'_3 = \frac{7 \cdot 637}{3 \cdot 363} \cdot 0,012456 = 0,051002;$$

$$f'_4 = \frac{6 \cdot 636}{4 \cdot 364} \cdot 0,051002 = 0,13367;$$

$$f'_5 = \frac{5 \cdot 635}{5 \cdot 365} \cdot 0,13367 = 0,23255;$$

$$f'_6 = \frac{4 \cdot 634}{6 \cdot 366} \cdot 0,23255 = 0,26855;$$

$$f'_7 = \frac{3 \cdot 633}{7 \cdot 367} \cdot 0,26855 = 0,19179;$$

$$f'_8 = \frac{2 \cdot 632}{8 \cdot 368} \cdot 0,19179 = 0,082345;$$

$$f'_9 = \frac{1 \cdot 631}{9 \cdot 369} \cdot 0,082345 = 0,015645.$$

Все компоненты дробной части этой формулы изменяются закономерно: в числителе убывают на единицу с каждым шагом, а в знаменателе возрастают на единицу с каждым шагом, чем можно воспользоваться для проверки расчета или для его упрощения.

Средняя арифметическая гипергеометрического распределения: $M = np$, где M — средняя арифметическая; n — объем пробы; p — доля вариант с учитываемым признаком в выборке; по данным § 1.27 она равна 0,635, а $q = 1 - p = 1 - 0,63 = 0,37$; $M = 9 \cdot 0,635 = 5,7$.

Дисперсия:

$$\sigma^2 = Mq \frac{N-n}{N-1}, \quad (1.136)$$

где σ^2 — дисперсия; N — объем совокупности, остальные обозначения — в тексте, перед формулой (1.136).

$$\sigma^2 = 5,7 \cdot 0,37 \cdot \frac{1008-9}{1008-1} = 2,09.$$

Коэффициент вариации гипергеометрического распределения:

$$v = \sqrt{\frac{q(N-n)}{M(N-1)}}, \quad (1.137)$$

$$v = \sqrt{\frac{0,37(1008-9)}{5,7(1008-1)}} = 0,254 \text{ или } 25,4\%.$$

При вычислении меди и сигмы обычным способом, принятым для нормального распределения по данным § 1.27, коэффициент вариации равен:

$$v = \frac{\sigma}{M} = \frac{1,829}{5,7} = 0,321, \text{ или } 32,1\%.$$

т. е. на 7% больше полученного по формуле (1.43), что следует учитывать при расчетах степени вариабельности подобных данных.

§ 1.29. Распределение Пуассона

Распределение Пуассона, или распределение редких событий, близко к биномиальному и образуется тогда, когда p (доля наблюдаемого признака) очень мало, а q близко к 1.

Таким образом, распределение Пуассона характерно для каких-либо редких событий, вероятность (доля) появления которых очень мала. Так же как и при биномиальном распределении, эмпирические частоты распределения Пуассона являются числом одинаковых проб, имеющих ту или иную долю наблюдаемого признака. В распределении Пуассона средняя арифметическая примерно равна дисперсии: $M \approx \sigma^2$, что является основным его признаком.

Теоретические частоты распределения Пуассона вычисляют по формуле:

$$f' = \frac{M^x}{x!} N_n e^{-M}, \quad (1.138)$$

где f' — теоретические частоты распределения Пуассона, т. е. число проб, обладающих той или иной долей наблюдаемого признака; x — варианты, отдельные значения наблюдаемого признака; $x!$ — (икс-факториал) обозначает произведение ряда натуральных чисел, например: $3! = 1 \cdot 2 \cdot 3 = 6$; M — средняя арифметическая данного ряда; N_n — общее число проб; $e = 2,71828182$, основание натуральных логарифмов (см. табл. 22П).

Первый способ. Вычислим теоретические частоты распределения Пуассона для данных табл. 1.34, где приведены данные распределения проб корнеплодов моркови, заболевших естественно (в отличие от данных из § 1.27), не поврежденных механически и без искусственного заражения их фомозом. Объем каждой пробы был $n = 9$, однако больше 4 больных корнеплодов в пробах не встречалось, поэтому расчет выполнен при $n = 4$, $N_n = 122$.

1. По формуле (1.13) определим среднюю арифметическую для данного взвешенного ряда:

$$M = \frac{\sum xf}{N_n} = \frac{0.70 + 1.32 + 2.14 + 3.2 + 4.4}{122} = 0,672.$$

2. Возведем M в степень с показателем x :

$$0,672^0 = 1, \quad 0,672^1 = 0,672, \quad 0,672^2 = 0,4516 \text{ и т. д. (столбец 3).}$$

3. В столбце 4 запишем факториалы от x :

$$0! = 1, \quad 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24.$$

4. Числа столбца 3 поделим на числа столбца 4 и частные запишем в столбце 5.

5. Вычислим выражение:

$$N_n e^{-M} = 122 \cdot 2,718^{-0,672} = 62,43.$$

6. Числа столбца 5 перемножим на число 62,43; произведения запишем в столбце 6.

Таблица 1.34. Вычисление частот распределения Пуассона

Число пораженных корнеплодов в пробе, x	Частоты (число проб), f	$M^x = 0,672^x$		$\frac{M^x}{x!}$	$\frac{M^x}{x!} N e^{-M}$	f'
0	70	1,000	$0! = 1$	1,00	62,43	62
1	32	0,6720	1	0,6720	41,95	42
2	14	0,4516	2	0,2258	14,10	14
3	2	0,3035	6	0,0506	3,16	3
4	4	0,2039	24	0,0085	0,53	1
122					122,77	122

7. После округления чисел столбца 6 до целых получаем теоретические частоты f' распределения Пуассона. Оценка совпадения эмпирических частот с теоретическими по критерию хи-квадрат (изложение метода его расчета см. § 4.14) показывает, что различие между ними несущественное ($\chi^2 = 4,413$; при объединении классов 3 и 4, $\nu = 4 - 2 = 2$; по табл. 10П требуется $\chi^2 = 5,99$ на уровне значимости $W'_1 = 0,05$).

Таким образом, можно сделать вывод, что естественное заражение фомозом корнеплодов моркови, не повреждаемых механически искусственным путем, — явление редкое, но встречающееся с определенной частотой, которое можно моделировать распределением Пуассона. По первому способу можно вычислять отдельно теоретическую частоту для любой варианты распределения Пуассона по формуле (1.138).

Второй способ более удобен для расчетов, когда требуется вычислять все частоты данного распределения; в этом случае для вычисления последующей теоретической частоты необходимо знать величину предыдущей.

Рассмотрим рекуррентный алгоритм вычисления частот по второму способу на примере распределения числа семян в плодах яблони домашней (табл. 1.35).

1. Вычисляем среднюю арифметическую и сигму

$$M = \frac{2996}{446} = 6,7175;$$

$$\sigma^2 = \frac{\sum x^2 f - (\sum x f)^2 / N}{N - 1} = \frac{23214 - 2996^2 / 446}{446 - 1} = 6,9402; \quad \sigma = 2,6344.$$

Проверяем соответствие данного ряда распределению Пуассона: средняя близка к величине дисперсии. Это подтверждает также их сравнение по критерию Фишера: $F = 6,9402 / 6,7175 = 1,03$, что меньше критической величины, равной 1,56; следовательно, данное распределение можно аппроксимировать распределением Пуассона.

Таблица 1.35. Распределение числа семян в плодах яблони домашней

Число семян, x	Число плодов, f	xf	x^2f	f'
0	1	0	0	0,53953
1	6	6	6	3,6243
2	15	30	60	12,173
3	38	114	342	27,257
4	44	176	704	45,775
5	40	200	1000	61,499
6	59	354	2124	68,854
7	60	420	2940	66,076
8	67	536	4288	55,483
9	49	441	3969	41,411
10	38	380	3800	27,818
11	14	154	1694	16,988
12	11	132	1584	9,5097
13	3	39	507	4,914
14	1	14	196	2,3578
15	0	0	0	1,0559
$k = 16$	446	2996	23 214	445,34

2. Находим нулевой член распределения с помощью логарифмов или таблиц e^{-x} :

$$f'_0 = N \cdot e^{-M} = 446 \cdot 2,7183^{-6,7175} = 0,53953.$$

3. По рекуррентной формуле:

$$f'_i = \frac{M}{i} f'_{i-1} \quad (1.139)$$

вычисляем все остальные члены ряда теоретических частот:

$$f'_1 = \frac{6,7175}{1} \cdot 0,53953 = 3,6243;$$

$$f'_2 = \frac{6,7175}{2} \cdot 3,6243 = 12,173;$$

$$f'_3 = \frac{6,7175}{3} \cdot 12,173 = 27,257$$

и т. д.; все теоретические частоты приведены в столбце 5 табл. 1.35, их удовлетворительное соответствие эмпирическим частотам достаточно ясно из рис. 6.

Третий способ вычисления частот распределения Пуассона основан на использовании таблиц, где для разных величин M приведены соответствующие частоты. Например, на стр. 197 таблиц Я. Янко (1961) для $M = 6,7$ и числа классов от 0 до 20 даны частоты, умножая которые на наш объем совокупности $N = 446$, получим

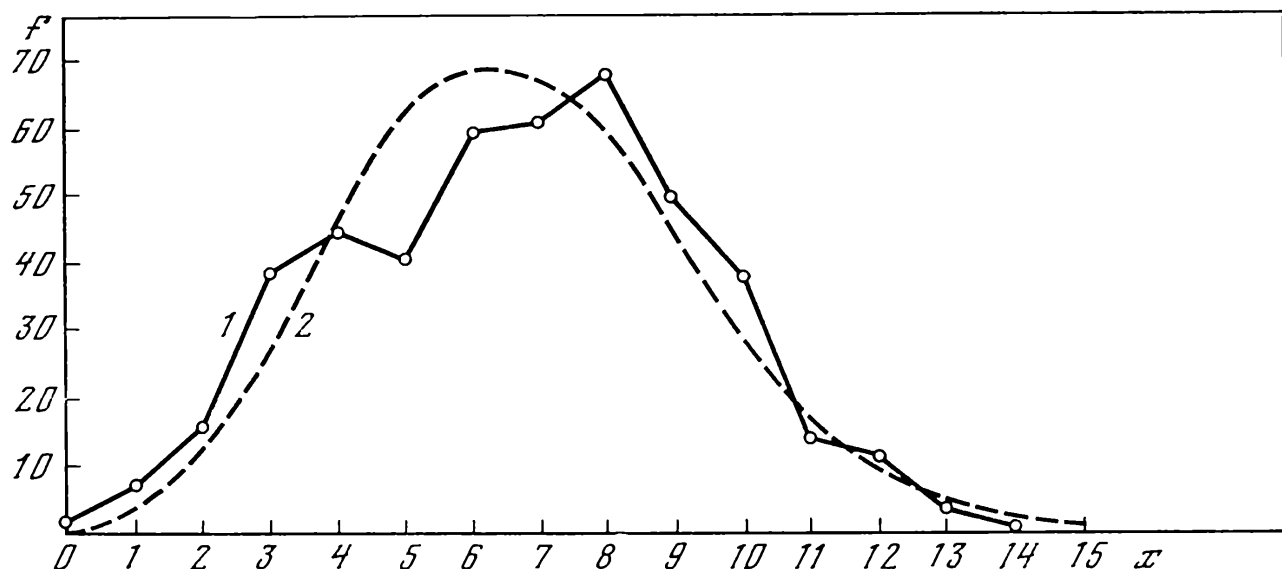


Рис. 6. Эмпирические (1) и теоретические (2) частоты (f) распределения Пуассона

x — число семян в плодах яблони

следующие частоты: 0,548; 3,68; 12,3; 27,5; 46,2; 61,6; 68,9; 66,0; 55,3; 41,2; 27,5; 16,8; 9,5; 4,82; 2,31; 1,03; 0,43; 0,169; 0,063; 0,022; 0,007, сумма которых 445,88. Соответствие этих частот вычисленным по первому и второму способам вполне удовлетворительное. Это позволяет рекомендовать третий способ для применения в первую очередь, поскольку для данного расчета необходима лишь логарифмическая линейка.

Центральные моменты, показатели асимметрии и эксцесса распределения Пуассона вычисляют по формулам:

$$\mu_2 = M, \quad \mu_3 = M, \quad \mu_4 = M(3M + 1), \quad (1.140)$$

$$A_s = \frac{1}{\sqrt{M}}, \quad E = \frac{1}{M} \quad (1.141)$$

Коэффициент вариации (v) в распределении Пуассона можно вычислить по формулам:

$$v = \frac{100}{\sigma} \%, \quad \text{или} \quad v = \frac{100}{\sqrt{M}} \%, \quad (1.142)$$

где σ — среднее квадратическое отклонение; M — средняя арифметическая.

Эти формулы можно использовать также для оценки близости некоторого эмпирического ряда к распределению Пуассона. Так, в примере, рассмотренном во втором и третьем способах, коэффициент вариации эмпирического ряда равен: $v = 100\sigma/M = 100 \cdot 2,6344/6,7175 = 39,2\%$, а по формуле (1.43) $v = 100/\sigma = 100/2,6344 = 38\%$. Обе величины достаточно близки друг к другу и, следовательно, данный эмпирический ряд соответствует распределению Пуассона. Более точно близость величин коэффициентов вариации можно оценить по формуле (4.07).

В литературе встречаются указания на возможность применения одного из свойств распределения Пуассона в качестве универсального критерия или правила для определения максимальных величин различных, в том числе биологических признаков. Речь идет о свойстве вариант этого распределения: $|x_i| < 2M$, где x_i — любая варианта совокупности; M — средняя арифметическая той же совокупности. На практике это, например, означает, что если для сортов некоторой культуры средняя урожайность равна M , то не может существовать сорта с урожайностью $x_i > 2M$. Фактически это правило представляет собой известное свойство более общего нормального распределения, в котором $2M = 6\sigma$, или $M = 3\sigma$, $x_i < 3\sigma = M$.

§ 1.30. Негативное биномиальное распределение

Распределение случайных, дискретных величин, вероятность появления которых в данной совокупности непостоянна, может быть моделировано негативным биномиальным распределением. На практике пользуются следующей формулой расчета членов этого распределения:

$$p^\alpha (1 - q)^{-\alpha}, \quad (1.143)$$

$$\alpha = \frac{M}{\sigma^2/M - 1}, \quad (1.144)$$

$$p = \frac{1}{1 + M/\alpha}, \quad (1.145)$$

$$q + p = 1, \quad (1.146)$$

где p — вероятность появления наблюдаемого признака; q — вероятность его отсутствия; α — экспонента бинома, характеризующая вариабельность данных; M , σ^2 — средняя и дисперсия эмпирического ряда.

В рассматриваемом типе распределения дисперсия по абсолютной величине должна быть больше средней арифметической той же совокупности; это служит одним из признаков того, что некоторый эмпирический ряд может быть моделирован негативным биномиальным распределением, вычисление частот которого ведется следующим образом. Определяются обычным путем величины M , σ^2 и способом моментов по формуле (1.144) — экспонента бинома. Дальнейшие вычисления обусловлены величиной полученной экспоненты α и ее соотношением с M , которое оценивается по формуле:

$$l = \frac{\Gamma(\alpha + M)(\alpha + 2)}{M} \geq 15. \quad (1.147)$$

Если $l \geq 15$, то расчеты ведем по первому способу. При $l < 15$, дополнительно вычисляем величину:

$$m = (M + 0,17)(f_1/N - 0,32) \geq 0,20, \quad (1.148)$$

где f_1 — первая эмпирическая частота ряда; N — объем совокупности.

Когда $m \geq 0,20$ и $l < 15$, расчеты ведем по второму способу. При $l < 15$, $m < 0,20$ переходим к третьему способу. Нередко оказывается, что, несмотря на несоответствие величин l и m критериям, указанным для первого способа, по нему получается неплохое совпадение на графике кривых эмпирических и теоретических частот. Поэтому в случаях, не требующих большой точности, можно ограничиться применением первого способа. Для эмпирических рядов, у которых $m \geq 0,20$ и $l < 15$, характерны резкая правосторонняя асимметрия кривых распределения и большая величина первой частоты по сравнению с остальными частотами данного ряда. Такие ряды обычно подходят для расчетов экспоненты бинома по второму способу.

Третий способ — универсальный, он дает точную величину экспоненты для любых форм негативного биномиального распределения, независимо от полученных величин l , m . Однако этот способ наиболее трудоемкий и рекомендуется для исполнения с помощью ЭВМ.]

Варианты (x) эмпирических рядов для вычислений по всем трем способам должны быть представлены числами 0, 1, 2, 3 и т. д. посредством преобразования: $x = (n_i - n_1)/c$ (см. примеры для второго и третьего способов). Три рассматриваемых способа отличаются друг от друга только техникой вычисления экспоненты α , в остальном их алгоритмы совпадают.

Первый способ. Расчеты параметров и частот рассмотрим на примере распределения числа побегов у двулетних растений гелениума осеннего (табл. 1.36). Средняя арифметическая, дисперсия, экспонента бинома и l -критерий для этого ряда равны:

$$M = 9,3397; \quad \sigma^2 = 19,71;$$

$$\alpha = 9,3397/[19,71/9,3397 - 1] = 8,4112;$$

$$l = \frac{(8,4112 + 9,3397)(8,4112 + 2)}{9,3397} = 19,787.$$

Так как $19,787 > 15$, то применяем к данному ряду первый способ, или способ моментов, как он иначе называется, т. е. величину α далее не уточняем и приступаем к общему для всех трех способов этапу работы — вычислению параметров p , q и теоретических частот. Находим по формулам (1.145), (1.146):

$$p = 1/(1 + 9,3397/8,4112) = 0,47385;$$

$$q = 1 - 0,47385 = 0,5261.$$

Первый член распределения (f_1) для $x = 0$ вычисляется особо по формуле:

$$f'_1 = p^n \text{ долей единицы, или } f'_1 = 100p^\alpha \%. \quad (1.149)$$

Для нашего примера с использованием логарифмов при любом их основании $f'_1 = 100 \cdot 0,47385^{8,4112} = \exp(\ln 100 + 8,4112 \cdot \ln 0,47385) = 0,1869\%$. Остальные члены разложения бинома в виде частот

**Таблица 1.36. Частоты негативного биномиального распределения
(первый способ)**

Порядковый номер i	Число побегов на двулетней особи гелениума осеннего, x	Частота, f	$f'\%$	f^e
1	0	0	0,1869	0,495
2	1	5	0,8272	2,192
3	2	12	2,048	5,427
4	3	9	3,739	9,908
5	4	14	5,611	14,87
6	5	14	7,327	19,42
7	6	23	8,616	22,83
8	7	13	9,332	24,73
9	8	33	9,458	25,06
10	9	22	9,073	24,05
11	10	17	8,311	22,02
12	11	14	7,318	19,39
13	12	26	6,228	16,5
14	13	17	5,144	13,63
15	14	13	4,139	10,97
16	15	9	3,253	8,621
17	16	6	2,504	6,637
18	17	5	1,892	5,014
19	18	6	1,405	3,724
20	19	2	1,028	2,723
21	20	4	0,741	1,964
22	21	1	0,5274	1,398
23	22	0	0,3709	0,983
$c = 1$	$k = 23$	265	99,08	262,6

стей в процентах" находим по рекуррентной формуле, т. е. такой, в которой каждый последующий член вычисляется по величине предыдущего:

$$f'_i = f'_{i-1} (\alpha + x_{i-1}) \frac{q}{x_i}, \quad (1.150)$$

т. е.

$$f'_2 = 0,1869 (8,4112 + 0) \frac{0,5261}{1} = 0,8272;$$

$$f'_3 = 0,8272 (8,4112 + 1) \frac{0,5261}{2} = 2,048;$$

$$f'_4 = 2,048 (8,4112 + 2) \frac{0,5261}{3} = 3,739 \text{ и т. д.}$$

При каждом новом шаге вычислений увеличиваются на единицу правый член выражения в скобках и знаменатель дроби. Теоретические частоты в процентах для всех 23 вариант ряда записаны

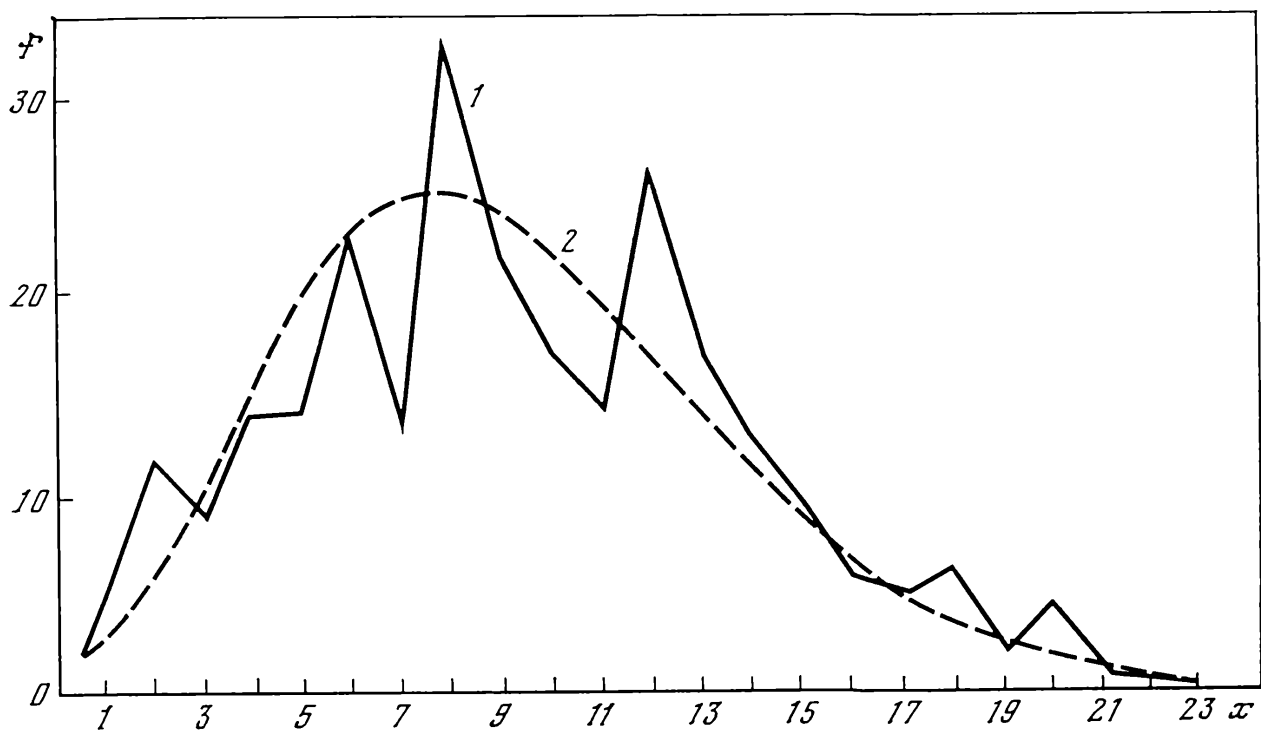


Рис. 7. Негативное биномиальное распределение] числа побегов (x) на двулетних растениях гелениума осеннего

f — частоты: 1 — эмпирические, 2 — теоретические

в столбце 4 табл. 1.36. Если перенести в них запятую влево на 2 знака, т. е. разделить эти числа на 100, а затем умножить на объем выборки $N = 265$, то получим теоретические частоты, которые записаны в последнем столбце таблицы. Буквой i в формуле (1.150) и табл. 1.36 обозначены порядковые номера вариантов и соответствующих им частот, начиная с номера 1. Варианты эмпирического взвешенного ряда перед вычислением частот отрицательного биномиального распределения желательно привести к натуральному ряду чисел, который начинается с нуля. Кривые эмпирических и теоретических частот в рассмотренном примере достаточно хорошо совпадают (рис. 7), что подтверждает и величина критерия согласия Колмогорова $K(\lambda) = 0,884$, которая меньше первого критического уровня, равного 1,36.

Таким образом, распределение числа побегов на одной двулетней особи у гелениума осеннего может быть моделировано отрицательным биномиальным распределением.

Второй способ, или способ пропорции первого члена. Рассмотрим технику вычислений по этому способу на примере распределения числа побегов на однолетних растениях гелениума осеннего (табл. 1.37). Вначале преобразуем исходные варианты ряда, приведенные в столбце 1 таблицы, в натуральный ряд, начинающийся с нуля (столбец 2), для чего из всех вариантов вычтем первую варианту и разделим их на классовой интервал $s = 1$, т. е. произведем действия по формуле, указанной в заголовке столбца 3 таблицы. В столбце 3 даны эмпирические частоты ряда, сумма которых равна объему совокупности $N = 339$. Вычислим для этих преобразован-

Таблица 1.37. Частоты негативного биномиального распределения (второй способ)

Число побегов на однолетних особях гелениума осеннего, n_i	$x = \frac{n_i - n_1}{c}$	f	$f\%$	f'
1	0	310	91,44	310
2	1	13	4,41	14,95
3	2	7	1,765	5,985
4	3	2	0,9138	3,098
5	4	3	0,5266	1,785
6	5	3	0,322	1,092
7	6	1	0,2045	0,6931
	7	0	0,1333	0,4517
	8	0	0,08852	0,3001
	9	0	0,05967	0,2023
	10	0	0,04069	0,1379
$c = 1;$	$k = 11$	339	99,9	338,7

ных вариант (x) следующие показатели: среднюю $M = 0,19469$; дисперсию $\sigma^2 = 0,60695$; экспоненту по формуле (1.144):

$$\alpha = \frac{0,19469}{0,60695/0,19469 - 1} = 0,091942,$$

критерий по формуле (1.147):

$$l = \frac{(0,091942 + 0,19469)(0,091942 + 2)}{0,19469} = 3,0799.$$

Так как вычисленная величина критерия меньше 15 ($l = 3,0799 < 15$) и поэтому данный вариационный ряд не желательно обрабатывать по первому способу, то определяем величину следующего критерия по формуле (1.148):

$$m = (0,19469 + 0,17) \left(\frac{310}{339} - 0,32 \right) = 0,22011.$$

Величина критерия m оказалась больше критической 0,20, поэтому экспоненту бинома для рассматриваемого эмпирического ряда следует вычислять по второму способу, т. е. по формуле:

$$\alpha_1 = \ln \frac{N}{f_1} / \left[\ln \left(1 + \frac{M}{\alpha_1} \right) \right], \quad (1.151)$$

где α_1 — экспонента бинома, полученная по второму способу; N — объем совокупности; f_1 — первая частота ряда, соответствующая первой варианте: $x = 0$. В результате вычисления по формуле (1.151) следует найти такое значение α_1 , при котором левая часть формулы будет равна правой. Алгоритм заключается в том, что вначале в правую часть формулы подставляют α_1 , найденное по этой же формуле (1.151). Полученный результат снова подстав-

ляют в правую часть, в знаменатель дроби, находящейся в скобках, и повторяют расчет до тех пор, пока подставленная и полученная величины α_1 не будут равны между собой с точностью до трех-четырех знаков после запятой. Равенство (1.151) в нашем примере соблюдается при $\alpha_1 = 0,064119$, а именно:

$$0,064119 = \ln \left(\frac{339}{310} \right) / \left[\ln \left(1 + \frac{0,19469}{0,064119} \right) \right]$$

Логарифмы в данном случае можно применять любые — натуральные или десятичные. Далее, по найденной таким образом величине α_1 вычисляем по формулам (1.145), (1.146):

$$p = \frac{1}{1 + 0,19469/0,064119} = 0,24774;$$

$$q = 1 - 0,24774 = 0,7523.$$

Начальный член равен (1.149):

$$f'_1 = 100 \cdot 0,24774^{0,064119} = 91,44\%.$$

Прочие члены находим по формуле (1.150):

$$f'_2 = 91,44 (0,064119 + 0) \frac{0,7523}{1} = 4,41\%;$$

$$f'_3 = 4,41 (0,064119 + 1) \frac{0,7523}{2} = 1,765\%;$$

$$f'_4 = 1,765 (0,064119 + 2) \frac{0,7523}{3} = 0,9135\% \text{ и т. д.}$$

В столбце 4 табл. 1.37 приведены теоретические частоты негативного биномиального распределения в процентах. Умножив их на объем выборки $N = 339$ и разделив на 100, получим теоретические частоты (столбец 5), которые достаточно хорошо совпадают с эмпирическими. Это подтверждает также и величина критерия согласия Колмогорова: $K(\lambda) = 0,111$. Заметим, что подобные распределения иногда можно аппроксимировать кривой Парето (§ 1.35).

Третий способ (максимального правдоподобия). Рассмотрим технику вычислений по этому способу на примере распределения числа детки у 347 сортов гладиолуса. В столбце 1 табл. 1.38 даны границы классов, по которым распределены данные о числе детки на одну луковицу, в столбце 2 — середины классов (n_i), в 3 — величины середин классов (x), преобразованные по формуле, данной в заголовке этого столбца, в столбце 4 — эмпирические частоты (f). По величинам x и f вычисляем, как обычно, среднюю арифметическую $M = 1,634$; дисперсию $\sigma^2 = 3,5391$ по формулам (1.144), (1.147):

$$\alpha = \frac{1,634}{3,5391/1,634 - 1} = 1,4015;$$

$$l = \frac{(1,4015 + 1,634)(1,4015 + 2)}{1,634} = 6,319.$$

Таблица 1.38. Частоты негативного биномиального распределения (третий способ)

Число детки на 1 луковицу гладиолуса		Середина класса, $x_i = \frac{n_i - n_1}{c}$	f	A	$f\%$	f'
Границы класса	Число деток, n_i					
0—9	5	0	111	236	33,58	116,53
10—19	15	1	103	133	25,67	89,06
20—29	25	2	58	75	16,64	57,74
30—39	35	3	23	52	10,14	35,2
40—49	45	4	22	30	5,988	20,78
50—59	55	5	11	19	3,465	12,03
60—69	65	6	8	11	1,979	6,866
70—79	75	7	4	7	1,119	3,882
80—89	85	8	3	4	0,628	2,179
90—99	95	9	4		0,3505	1,216
		10			0,1947	0,6755
		11			0,1077	0,3739
$c = 10$	$k = 10$	—	347	—	99,86	346,53

Так как вычисленная величина числа l меньше 15, находим следующий критерий (1.148):

$$m = (1,634 + 0,17) \left(\frac{111}{347} - 0,32 \right) = 0,00284,$$

полученная величина которого значительно меньше 0,20. Поэтому к данному эмпирическому ряду следует применить третий способ для уточнения величины экспоненты бинома, основанный на решении уравнения:

$$\sum_i \left(\frac{A_i}{x + \alpha_2} \right) = N \ln \left(1 + \frac{M}{\alpha_2} \right), \quad (1.152)$$

где A — накопленные частоты; x — варианты ряда из столбца 3 табл. 1.38; α_2 — уточненное значение экспоненты бинома; N — объем совокупности; M — средняя арифметическая.

Уравнение (1.152) можно решать путем итераций, или последовательных приближений, на ЭВМ или с помощью ручных средств счета с применением интерполяции для ускорения приближения к равенству обеих частей уравнения.

В столбце 5 таблицы приведены накопленные частоты, которые вычисляются путем последовательного вычитания всех эмпирических частот из объема совокупности, в данном случае из $N = 347$. Например, $A_1 = 347 - 111 = 236$; $A_2 = 236 - 103 = 133$; $A_3 = 133 - 58 = 75$; $A_4 = 75 - 23 = 52$ и т. д. до $k - 1$.

Для данных рассматриваемого примера уравнение (1.152) спра-

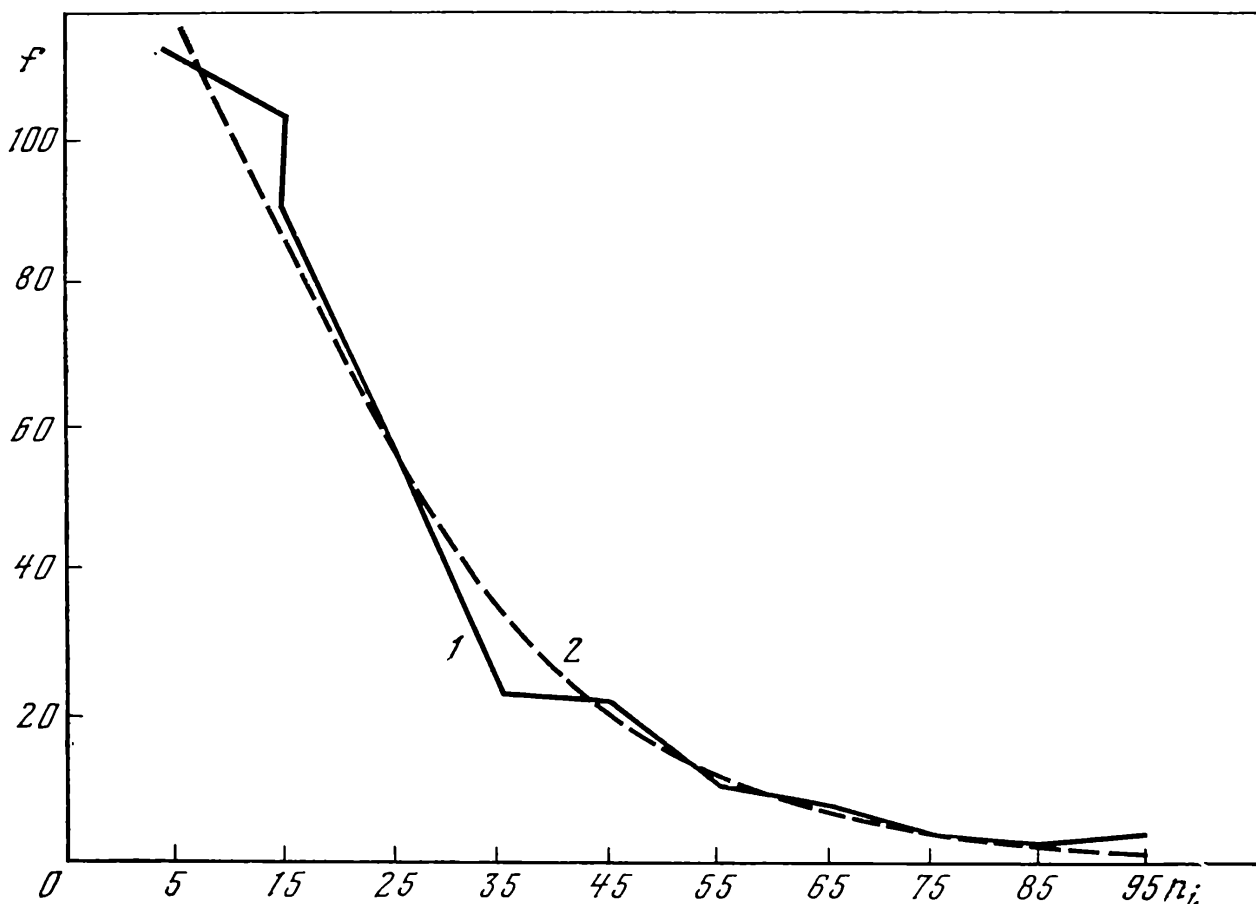


Рис. 8. Негативное биномиальное распределение числа детки (n) у луковиц гладиолуса

f — частоты: 1 — эмпирические, 2 — теоретические

ведливо при $\alpha_2 = 1,4359$, т. е.

$$\begin{aligned} & \frac{236}{0 + 1,4359} + \frac{133}{1 + 1,4359} + \frac{75}{2 + 1,4359} + \frac{52}{3 + 1,4359} + \\ & + \frac{30}{4 + 1,4359} + \frac{19}{5 + 1,4359} + \frac{11}{6 + 1,4359} + \frac{7}{7 + 1,4359} + \\ & + \frac{4}{8 + 1,4359} = 347 \ln \left(1 + \frac{1,634}{1,4359} \right). \end{aligned}$$

За исходное значение экспоненты перед началом процесса итерации принимаем то, которое было получено по формуле (1.144), в данном случае $\alpha = 1,4015$. Изменяя его в ту или другую сторону, добиваемся, чтобы выполнялось равенство (1.152). После вычисления α_2 определяем величины (1.145), (1.146):

$$p = \frac{1}{1 + \frac{1,634}{1,4359}} = 0,46773; \quad q = 1 - 0,46773 = 0,5323.$$

Начальная теоретическая частота (1.149):

$$\begin{aligned} f'_1 &= 100 \cdot 0,46773^{1,4359} = \exp(\ln 100 + 1,4359 \ln 0,46773) = \\ &= 33,58\%. \end{aligned}$$

Остальные частоты (столбец 6 табл. 1.38) вычисляем по рекур-

рентной формуле (1.150):

$$f'_2 = 33,58 (1,4359 + 0) \frac{0,5323}{1} = 25,67 \%,$$

$$f'_3 = 25,67 (1,4359 + 1) \frac{0,5323}{2} = 16,64\% \text{ и т. д.}$$

(до 12 класса). Переноса в них запятую на два знака влево и умножая на объем совокупности $N = 347$, получим теоретические частоты (столбец 7), которые хорошо соответствуют эмпирическим частотам, как видно из рис. 8 и величины критерия согласия Колмогорова $K(\lambda) = 0,4655$. Число классов теоретических частот и частот в данном случае на два класса больше числа эмпирических ввиду того, что суммы столбцов 6 и 7 иначе были бы заметно меньше соответственно 100% и $N = 347$.

§ 1.31. Мультиномиальное распределение

Все до сих пор рассмотренные здесь дискретные распределения образованы изменением одной случайной величины и другой, альтернативно связанной с ней. При расширении числа комбинаций признаков, с определенной вероятностью независимо друг от друга появляющихся в генеральной совокупности, получается мультиномиальное распределение.

Вероятность появления определенной комбинации признаков, или k -мерной случайной величины, вычисляется по следующей формуле мультиномиального распределения:

$$P = \frac{N! p_1^{n_1} \cdot p_2^{n_2} \cdot \dots \cdot p_k^{n_k}}{n_1! n_2! \cdot \dots \cdot n_k!}, \quad (1.153)$$

где P — вероятность появления комбинации из k признаков; n_1, n_2, \dots, n_k — эмпирические численности особей с данными признаками; $N = n_1 + n_2 + \dots + n_k$ — объем пробы, взятой из генеральной совокупности; $p_1 - p_k$ — теоретические вероятности появления признаков.

В биологии мультиномиальное распределение, в частности, может быть применено для оценки степени случайности появления некоторых комбинаций признаков в генетических, фитосоциологических и других исследованиях.

Например, в одной коллекции сортов картофеля насчитывалось 209 сортов с белой окраской венчиков цветков, 88 — с красно-фиолетовой, 56 — с красной, 24 — с темно-сине-фиолетовой и 20 сортов с голубой окраской венчиков, всего 397 сортов. Доли этих окрасок, или признаков, соответственно составляют: $p_1 = 0,526$; $p_2 = 0,222$; $p_3 = 0,141$; $p_4 = 0,06$; $p_5 = 0,05$ от общего числа сортов. В пробе, взятой случайным образом из популяции гибридных семян, полученных в результате свободного опыления, насчитали $n_1 = 8$ семян с белой окраской венчиков, $n_2 = 6$ с красно-фиолетовой, $n_3 = 4$ с красной, $n_4 = 6$ с темно-сине-фиолетовой, $n_5 = 5$ с голубой, всего $N = 29$ семян.

По формуле (1.153):

$$P = \frac{29! 0,526^8 \cdot 0,222^6 \cdot 0,141^4 \cdot 0,06^6 \cdot 0,05^5}{8! 6! 4! 6! 5!},$$

применяя десятичные логарифмы и таблицы логарифмов факториалов [например, (Большев, Смирнов, 1968, с. 454)], найдем искомую вероятность:

$$\begin{aligned} \lg P &= 30,94654 - (4,60552 + 2,85733 + 1,38021 + 2,85733 + \\ &+ 2,07918) - 8 \cdot 0,279 - 6 \cdot 0,6536 - 4 \cdot 0,8508 - \\ &- 6 \cdot 1,2218 - 5 \cdot 1,301 = -6,22563; \lg P = \bar{7},77437, \\ &\text{откуда } P = 0,000000889. \end{aligned}$$

Таким образом, вероятность случайного появления данной комбинации численностей очень мала: из 10 млн. проб, по 29 особей в каждой, может встретиться примерно лишь 9 проб, в которых будет наблюдаться указанное числовое соотношение признаков окраски цветков, появление которого было, следовательно, вызвано в данном случае некоторыми существенно действующими факторами. Формулу (1.153) можно также применить для анализа расщепления гибридного потомства, подобно тому, как это сделано далее в § 4.14, посредством критерия хи-квадрат.

§ 1.32. Распределение Максвелла

Кривая распределения эмпирических частот вариационного ряда существенно положительных чисел, обладающего умеренной асимметрией, может быть аппроксимирована распределением Максвелла. Кривая частот рассматриваемого типа распределения удобна тем, что строится посредством уравнения, определяемого всего одним параметром. Она достаточно гибка и может быть применена для выравнивания асимметричных распределений различных биологических признаков, величины которых изменяются по непрерывному типу.

Рассмотрим технику вычисления частот распределения Максвелла на примере эмпирического распределения периодов продолжительности цветения в днях 381 вида деревьев и кустарников в Ленинграде. Этот ряд, приведенный в столбцах 1 и 2 табл. 1.39, удовлетворяет требованиям, перечисленным в начале параграфа: период цветения — существенно положительная величина и не может принимать отрицательных значений; ряд — умеренно асимметричен (рис. 9, 1). Последовательность расчетов следующая.

1. Вычисляем среднюю арифметическую: $M = 16,363$ и среднее квадратическое отклонение ряда $\sigma = 12,362$ для оценки теоретического соответствия условиям аппроксимации посредством коэффициента вариации (v) распределения Максвелла, который является величиной постоянной и равен:

$$v = \sqrt{\frac{3\pi}{8}} - 1 = 0,42203, \text{ или } 42\%. \quad (1.154)$$

Таблица 1.39. Вычисление частот распределения Максвелла:

Период цветения (в днях) видов деревьев и кустар- ников, x	Число видов, f	$tz = \left(\frac{x}{\alpha}\right)^2$	$e^{-0,5tz}$	$tz e^{-0,5tz}$	f'_0	f'
1	2	3	4	5	6	7
5	37	0,23777	0,88798	0,21114	0,16429	62,595
15	238	2,1401	0,34298	0,73401	0,57115	217,61
25	78	5,9443	0,051196	0,30432	0,2368	90,221
35	9	11,651	2,9514·10 ⁻³	3,4387·10 ⁻²	0,026757	10,194
45	8	19,259	6,5764·10 ⁻⁵	1,2666·10 ⁻³	9,8557·10 ⁻⁴	0,3755
55	3	28,77	5,6593·10 ⁻⁷	1,6282·10 ⁻⁵	1,2669·10 ⁻⁵	4,8269·10 ⁻³
65	3	40,183	1,8803·10 ⁻⁹	7,5556·10 ⁻⁸	5,8792·10 ⁻⁸	2,24·10 ⁻⁵
75	1	53,498	0	0	0	0
85	2	68,716	0	0	0	0
95	0	85,835	0	0	0	0
105	1	104,86	0	0	0	0
115	1	125,78	0	0	0	0
$c = 10; \quad k = 12$						381
						1,00000
						381,00

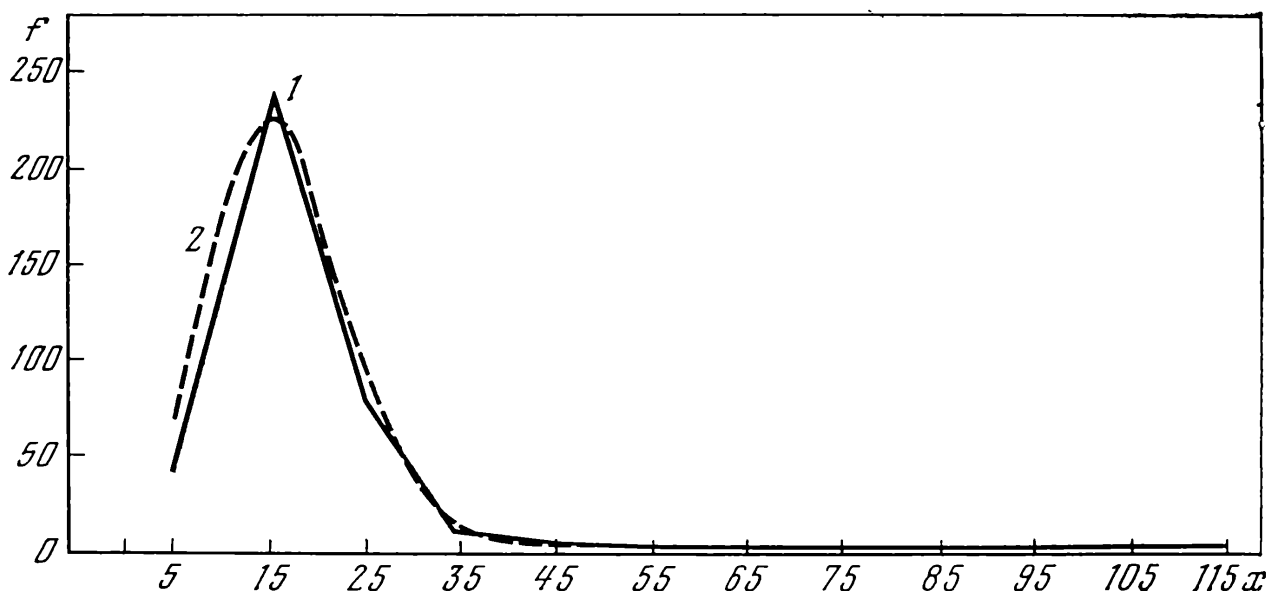


Рис. 9. Распределение Максвелла периодов цветения (x) древесных растений в Ленинграде

f — частоты: 1 — эмпирические, 2 — теоретические

Отклонения от данного соотношения, однако, могут быть довольно большими, в частности, для рассматриваемого примера (1.43): $v = \frac{\sigma}{M} = \frac{12,362}{16,363} = 0,75$, или 75%, тем не менее качество аппроксимации данного ряда распределением Максвелла, как увидим дальше, нельзя назвать плохим.

2. Параметр распределения получаем по одной из четырех равноценных формул:

$$\alpha = \frac{M}{2\sqrt{2/\pi}}, \quad \alpha = \frac{M}{\sqrt{8/\pi}}, \quad \alpha = \frac{M}{1,595769}, \quad \alpha = 0,626657M, \quad (1.155)$$

где M — средняя арифметическая; α — параметр распределения, т. е. в нашем примере:

$$\alpha = \frac{16,363}{2\sqrt{2/\pi}} = 10,254.$$

3. Среднее квадратическое отклонение распределения Максвелла:

$$\sigma = \alpha \sqrt{3 - \frac{8}{\pi}}; \quad (1.156)$$

$$\sigma = 10,254 \cdot 0,67345 = 6,9053, \quad \sigma^2 = 47,683.$$

4. Мода:

$$M_0 = \alpha \sqrt{2}; \quad (1.157)$$

$$M_0 = 10,254 \cdot 1,41421 = 14,501.$$

5. Медиана:

$$M_e = 1,538\alpha; \quad (1.158)$$

$$M_e = 1,538 \cdot 10,254 = 15,771.$$

В случае асимметричных распределений мода и особенно медиана нередко лучше характеризуют ряд по его центральной тенденции, чем средняя арифметическая.

6. В столбце 3 табл. 1.39 подсчитаны квадраты нормированных отклонений, для чего середины интервалов x поделены на $\alpha = 10,254$ и возведены в квадрат.

7. Числа столбца 3 умножены на $-0,5$ и от них получены экспоненты, которые легко найти по таблицам e^x [например, (Сегал, Семендяев, 1962)], или путем логарифмирования; результаты записаны в столбце 4.

8. Числа столбца 4 умножены на числа из столбца 3 и результаты записаны в столбце 5.

9. Константа $\frac{c}{\alpha} \sqrt{\frac{2}{\pi}} = \frac{10 \cdot 0,797885}{10,254} = 0,77812$, где c — классовый интервал, умножена на числа столбца 5 и полученные частоты в долях единицы записаны в столбце 6. Если перенести запятую в этих числах на 2 знака вправо, получим частоты в процентах.

10. Умножая числа столбца 6 на объем выборки $N = 381$, получим теоретические частоты распределения Максвелла (столбец 7), которые на рис. 9, 2 достаточно хорошо совпадают с эмпирическими частотами [критерий Колмогорова $K(\lambda) = 1,31$]. Изложенный здесь и рекомендуемый для применения ход вычислений соответствует формуле:

$$f' = \frac{Nc}{\alpha} \sqrt{\frac{2}{\pi}} t^2 e^{-0,5t^2}, \quad (1.159)$$

где f' — теоретические частоты распределения Максвелла; N — объем выборки; α — параметр распределения; $t = x/\alpha$ — нормированное отклонение; x — середины интервалов эмпирического ряда; e — основание натуральных логарифмов.

Формула (1.159) может быть представлена также и в следующих двух вариантах:

$$f' = \frac{Ncx^2}{\alpha^3} \sqrt{\frac{2}{\pi}} \cdot e^{-\frac{x^2}{2\alpha^2}} \quad f' = \frac{2Nc}{\sqrt{2\pi}} \cdot \frac{t^2}{\alpha} e^{-\frac{t^2}{2}},$$

где обозначения те же. При вычислении параметров и частот распределения Максвелла полезно располагать значениями следующих констант:

$$\pi = 3,14159; \quad e = 2,718282; \quad 2 \sqrt{\frac{2}{\pi}} = 1,595769;$$

$$\frac{1}{2 \sqrt{\frac{2}{\pi}}} = 0,626657; \quad \sqrt{\frac{2}{\pi}} = 0,797885; \quad \sqrt{2} = 1,414214;$$

$$\frac{1}{\sqrt{2\pi}} = 0,398942; \quad 3 - \frac{8}{\pi} = 0,453531; \quad \sqrt{3 - \frac{8}{\pi}} = 0,67345.$$

Если интервалы ряда после его составления окажутся слишком крупными, как это получилось в ряду из табл. 1.39, то промежуточные частоты можно получить уменьшая классовой интервал и одновременно в столько же раз увеличивая объем выборки и число классов ряда. Уменьшив классовой интервал в 2 раза ($c = 5$) и увеличив объем выборки в 2 раза ($N = 762$) и число классов тоже в 2 раза ($k = 24$), получим по формуле (1.159) для соответствующих вариантов следующие частоты в начале кривой: 0—0; 5—62,6; 10—175,3; 15—217,6; 20—168,3; 25—90,2; 30—35,1; 35—10,2; 40—2,2, которые помогают более точно построить кривую теоретического распределения на том участке, где она была недостаточно определена из-за малого числа точек. Распределение Максвелла связано с трехмерным нормальным распределением, поэтому его частоты можно вычислить также при помощи двух таблиц по формуле:

$$f_{\Sigma} = N [F(t) - 2t\psi(t)], \quad (1.160)$$

где f_{Σ} — накопленные частоты распределения Максвелла; N — объем совокупности; $F(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$ — интегральная функция нормального распределения; $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ — плотность нор-

мального распределения; $t = x/\alpha$ — нормированное отклонение; x — варианты эмпирического распределения (середины классов); α — параметр распределения Максвелла.

Вычисления по второму способу ведутся следующим образом.

1. По величинам t : 0,48761; 1,4628 и т. д. из таблицы $F(t)$ выбираются соответствующие числа [они даны, например, в работе А. К. Митропольского (1952)]. Для первых двух вариантов они равны: 0,37445; 0,856533.

2. Из табл. 5П по тем же величинам t находим (с интерполяцией) значения функции $\psi(t)$: 0,24187; 0,01614; умножаем их на $2t$: 0,9752; 2,9256; вычитаем из найденных $F(t)$, получаем: 0,13258; 0,84039; умножаем на $N = 381$ и получаем накопленные частоты 50,5; 320 и таким же образом еще три: 372, 380, 381.

3. Дальнейшее вычисление накопленных частот прекращаем, так как все последующие будут равны объему выборки 381. Вычитая предыдущую частоту из последующей, получаем 12 частот: 50,5; 269,5; 52; 28; 1; 0 (остальные 6 — нули).

Как видим, найденные по второму способу частоты, хотя и достаточно близки к эмпирическим, но отличаются от найденных по первому способу. Объясняется это тем, что частоты данного эмпирического ряда сильно концентрированы в начале кривой, которая к тому же имеет диапазон вариации от нуля до $11t$, на что существующие таблицы $F(t)$ и $\psi(t)$ не рассчитаны. Из этого следует, что в случае сильной концентрации частот в основной части ряда и большого варьирования дат лучше применять первый способ вычислений, и напротив, при более мелких интервалах ряда и сравнительно

небольшого диапазона вариации (до $4t$ в обе стороны) можно применять второй способ. Первый способ, кроме того, как показал опыт, удобнее для программирования на ЭВМ.

Из приведенного анализа можно сделать следующие выводы. Варьирование периода цветения у видов деревьев и кустарников очень велико. Величина нормы для периода цветения этих видов должна находиться в пределах $M \pm \sigma = 16,363 \pm 12,362$, т. е. от 4 до 29 дней. В этот период укладывается цветение 362 видов из 381, т. е. подавляющее большинство. Остальные 19 видов относятся или к ремонтантным, или при наблюдениях ошибочно фиксировались (главным образом, в конце лета и осенью) вторично цветущие виды.

Таким образом, критерий нормальной продолжительности цветения (4—29 дней) дает способ выделить отклоняющиеся виды для последующего биологического анализа их фенологии. В частности, этот критерий позволяет количественно отделить ремонтантные виды от видов с нормальной продолжительностью цветения. Большинство видов (238) цветет в течение 15 дней, на что указывает величина моды. Половина видов цветет в течение менее 16 дней, другая половина — дольше этого периода, что видно из величины медианы.

§ 1.33. Распределение Рэля

К числу типов распределений, определяемых одним параметром, относится также распределение Рэля, которое пригодно для аппроксимации умеренно асимметричных вариационных рядов, составленных из дат непрерывного вида, принимающих только положительные значения. Количественным критерием того, что некоторый эмпирический взвешенный ряд может быть аппроксимирован распределением Рэля, служит соотношение:

$$\sigma = M \sqrt{\frac{4}{\pi}} - 1 = 0,5227M, \quad (1.161)$$

т. е. среднее квадратическое отклонение должно быть примерно в 2 раза меньше средней арифметической.

Например, распределение 205 сортов пиона по среднему из 10 особей числу побегов (фертильных и вегетативных вместе) на особи может состоять только из положительных величин, которые после усреднения изменяются по непрерывному типу. Показатели этого эмпирического ряда, приведенного в табл. 1.40 (столбцы 1—3), следующие: $M = 22,075$; $\sigma = 11,7806$; $A_s = 0,60436$; $E = -0,0842$.

Вычисление показателей и частот распределения Рэля ведется в следующем порядке.

1. Основной параметр (a) найдем по формуле:

$$a^2 = \frac{2M^2}{\pi}, \quad (1.162)$$

где M — средняя арифметическая эмпирического ряда; $\pi = 3,1416$,

$$a^2 = \frac{2 \cdot 22,075}{3,1416} = 310,24.$$

Таблица 1.40. Частоты распределения Рэля

Общее число побегов на особи пиона		Частота, f	$\frac{x}{a^2}$	$\frac{x}{a^2} e^{-\frac{x^2}{2a^2}}$	f_0'	f'
Границы интервала	Середина интервала, x					
1,5—8,4	5	21	0,016117	0,015479	0,10835	22,212
8,5—15,4	12	48	0,03868	0,030667	0,21467	44,007
15,5—22,4	19	50	0,061243	0,034229	0,2396	49,118
22,5—29,4	26	29	0,083806	0,02819	0,19733	40,453
29,5—36,4	33	34	0,10637	0,01839	0,12873	26,39
36,5—43,4	40	13	0,12893	0,0097821	0,068475	14,037
43,5—50,4	47	7	0,1515	0,0043078	0,030155	6,1818
50,5—57,4	54	2	0,17406	0,0015839	0,011087	2,2728
57,5—64,4	61	1	0,19662	$4,889 \cdot 10^{-4}$	0,0034223	0,70157
$= 7; k = 12$		205			1,0018	205,37

2. Мода равна: $M_0 = a = \sqrt{310,24} = 17,614$.

3. Дисперсия:

$$\sigma^2 = a^2 \left(2 - \frac{\pi}{2} \right), \quad (1.163)$$

$$\sigma^2 = 310,24 \left(2 - \frac{3,1416}{2} \right) = 133,16.$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\sigma^2} = \sqrt{133,16} = 11,54.$$

4. Медиана:

$$M_e = \sqrt{2a^2 \ln 2}; \quad (1.164)$$

$$M_e = \sqrt{2 \cdot 310,24 \cdot 0,69315} = 20,738.$$

5. Частоты вычисляем по формуле:

$$f' = \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} \cdot Nc, \quad (1.165)$$

где f' — частоты распределения Рэля; x — середины классов эмпирического распределения; a — основной параметр; $e = 2,7183$; N — объем выборки; c — классовой интервал.

$$f' = \frac{x}{310,24} e^{-\frac{x^2}{620,48}} 205 \cdot 7.$$

В табл. 1.40, столбце 4 приведены частные от деления середин классов на a^2 :

$$5/310,24 = 0,016117 \text{ и т. д.}$$

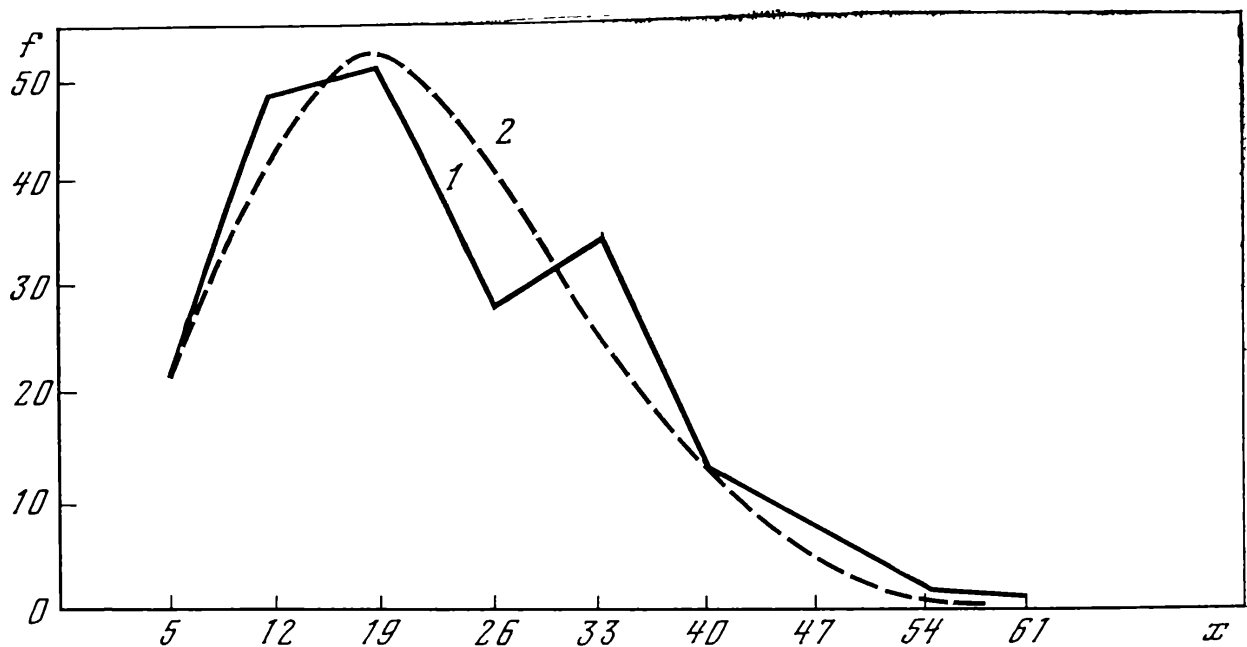


Рис. 10. Распределение Рэля числа побегов (x) на растениях пиона
 f' — частоты: 1 — эмпирические, 2 — теоретические

В столбце 5 эти величины умножаются на $e^{-\frac{x^2}{2a^2}}$,

$$(5/310,24) \cdot (e^{-\frac{5^2}{620,48}}) = 0,015479.$$

Числа столбца 5, умноженные на $c = 7$, дают теоретические частоты (столбец 6), перенося в которых запятую на 2 знака вправо, получим частоты в процентах. Умножив числа столбца 6 на $N = 205$, получим теоретические частоты распределения Рэля (столбец 7).

Критерий согласия Колмогорова оказался равным $K(\lambda) = 0,54$, что позволяет говорить о достаточно хорошем совпадении эмпирических и теоретических частот (рис. 10).

§ 1.34. Показательное распределение

По этому типу распределения выравнивают эмпирический ряд, частоты которого распределены с резко выраженной правосторонней (или положительной) асимметрией, а средняя арифметическая примерно равна сигме: $M \approx \sigma$. При этом величины изучаемого признака должны изменяться непрерывно и принимать только неотрицательные значения, т. е. они могут быть положительными числами, или равными нулю. Единственный параметр показательного распределения — также всегда число положительное, обратное средней арифметической; вычисляется он по формуле:

$$\lambda = \frac{1}{M}, \quad (1.166)$$

где λ — параметр показательного распределения; M — средняя арифметическая данного эмпирического ряда. Дисперсия, сигма и

медиана показательного распределения вычисляются по формулам:

$$\sigma^2 = \frac{1}{\lambda^2}, \quad \sigma = \frac{1}{\lambda}, \quad (1.167)$$

$$M_e = \frac{|\ln 2|}{\lambda} = \frac{|0,69315|}{\lambda} = 0,69315M, \quad (1.168)$$

где σ^2 — дисперсия; σ — сигма; λ — параметр распределения; M_e — медиана; M — средняя арифметическая. Коэффициент вариации, учитывая равенство средней и сигмы в этом распределении равен: $v = \sigma/M = 1$, или 100%, что может служить одним из признаков соответствия эмпирического ряда показательному распределению. Мода показательного распределения всегда равна нулю, или началу отсчета вариант. Теоретические частоты (f') вычисляют по формуле:

$$f' = N(e^{-\lambda g_{i-1}} - e^{-\lambda g_i}), \quad (1.169)$$

где λ — параметр распределения; g_{i-1} , g_i — верхние границы предыдущего и последующего классов распределения вариант изучаемой величины; N — объем совокупности.

Распределение 2619 побегов в зависимости от числа соцветий на них у однолетних растений гелениума осеннего, приведенное в столбцах 1—3 табл. 1.41, обладает значительной положительной асимметрией: $A_s = 1,4563$, эксцесс также велик: $E = 2,34$, коэффициент вариации $v = 83,4\%$, т. е. близок к 100%, величина средней арифметической $M = 44,828$ близка к величине сигмы $\sigma = 37,383$; число соцветий — величина существенно положительная, изменяющаяся в данном случае непрерывно, так как исходные данные являются результатом усреднения — все перечисленные признаки позволяют применить к этому ряду показательное распределение для выравнивания эмпирических частот. Несмотря на большое варьирование данных, показатель точности опыта сравнительно мал: $P = 1,63\%$ и находится в допустимых пределах ($P \leq 5\%$).

Кроме того, в табл. 1.41 приведены следующие данные, которые разъясняются ниже.

В соответствии с формулой (1.169) для расчетов использованы не середины классов, а их верхние (меньшие) границы, которые записаны в столбце 4 на одну строку выше, так как объем интервала вычисляется по разности его границ.

В столбце 5 приведена величина параметра распределения по формуле (1.166)

$$\lambda = \frac{1}{44,828} = 0,022307,$$

умноженная на все значения верхних границ из столбца 4.

Столбец 6 — экспоненты чисел из столбца 5, взятых с минусом перед ними.

В столбце 7 дана разность предыдущего и последующего чисел из столбца 6, например: $1 - 0,5121 = 0,4879$; $0,5121 - 0,26227 = 0,2498$ и т. д. Эти числа представляют собой вероятности появ-

Таблица 1.41. Вычисление частот показательного распределения

Границы класса	Сере- дина клас- са, x	Часто- та, f	Верхняя граница класса, g	λg	$e^{-\lambda g}$	$e^{-\lambda g_{i-1}} -$ $e^{-\lambda g_i}$	$N \cdot (7) = f'$	$f' \%$
1	2	3	4	5	6	7	8	9
			0	0	1	—	—	—
0—29	15	1249	30	0,66921	0,5121	0,4879	1277,8	48,8
30—59	45	639	60	1,3384	0,26227	0,2498	654,31	25
60—89	75	419	90	2,0076	0,1343	0,128	335,15	12,8
90—119	105	185	120	2,6768	0,06879	0,06551	171,57	6,56
120—149	135	89	150	3,3461	0,035222	0,03357	87,915	3,36
150—179	165	20	180	4,0153	0,018039	0,01718	45,002	1,72
180—209	195	10	210	4,6845	0,009239	0,0088	23,047	0,88
210—239	225	8	240	5,3537	0,0047305	0,004508	11,808	0,45
240—269	255	0	270	—	—	0,002308	6,0444	0,23
$k = 9 \quad c = 30 \quad 2619$						0,998	2612,6	99,8

ления побегов с данным количеством соцветий на них; сумма этих вероятностей должна быть равна 1.

Умножив полученные вероятности на объем совокупности $N = 2619$ побегов, получим теоретические частоты показательного распределения (столбец 8), которые в столбце 9 представлены также в процентах от общей их суммы: 2612,6. Частоты теперь соответствуют серединам классов из столбца 2. Как видим, сумма теоретических частот несколько меньше фактической; это происходит в подобных случаях в основном из-за неполноты теоретического ряда, вычисление частот которого можно было бы продолжить и получить недостающие 6,4 побегов, или 0,2% от объема исходной совокупности. Однако эти дальнейшие частоты становятся слишком малыми, маловероятными событиями, поэтому здесь не приводятся.

Сравнение фактических частот (столбец 3) с теоретическими (столбец 8) по критерию Колмогорова показало, что их совпадение вполне удовлетворительное: $K(\lambda) = 1,059$ (рис. 11).

Дисперсия и сигма показательного распределения на практике используются редко. Вместо них обычно применяют дисперсию и сигму, подсчитанные способом, обычным для нормального распределения. В частности, для рассматриваемой совокупности второй центральный момент равен 1397,5.

Медиана показательного распределения по формуле (1.168) равна

$$M_e = 0,69315 \cdot 44,828 = 31,072 \text{ соцветия.}$$

Отдельно взятые частоты показательного распределения, соответствующие серединам классов, вычисляются по формуле:

$$f' = Nc\lambda e^{-\lambda x}, \quad (1.170)$$

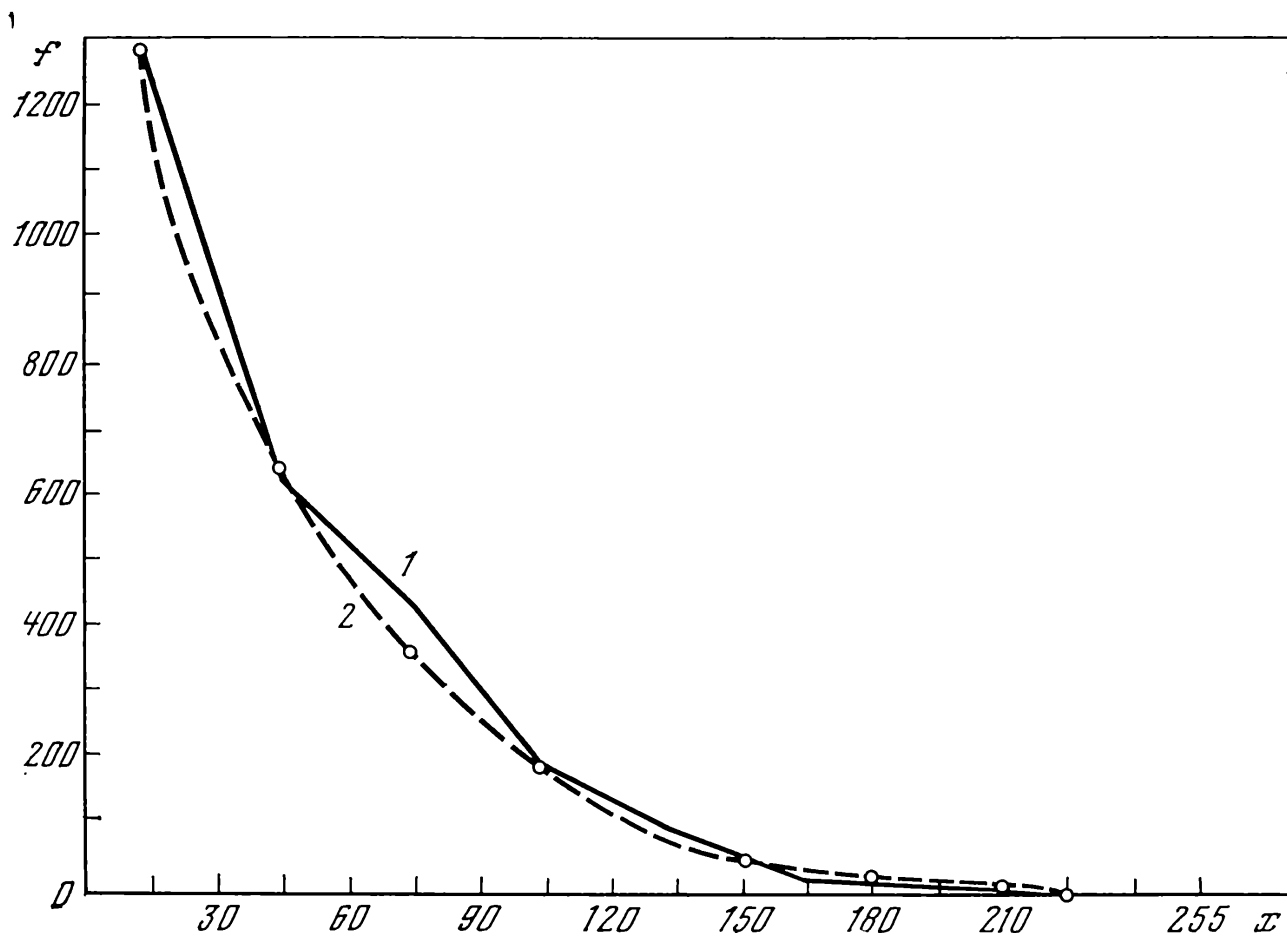


Рис. 11. Показательное распределение побегов (f) у однолетних растений гелениума осеннего в зависимости от числа соцветий на них (x)

f — частоты: 1 — эмпирические, 2 — теоретические

где x — середины классов; c — интервал распределения; остальные обозначения те же, что и к формуле (1.169).

Например, при $x = 15$ соцветий:

$$f' = 2619 \cdot 30 \cdot 0,0022307 e^{-0,0022307 \cdot 15} = 1254,2 \text{ побега.}$$

Заметим, что вытекающее из условия показательного распределения равенство $M = \sigma$ не рекомендуется улучшать за счет усреднения: $(M + \sigma)/2$ исходных показателей, так как это, по-нашему опыту, не приводит к улучшению аппроксимации кривой распределения, т. е. в формулу (1.166) и последующих расчетах следует подставлять только среднюю (M) исходного эмпирического ряда.

Если возвести в квадрат среднюю и сигму: $M^2 = (44,828)^2 = 2009,55$; $\sigma^2 = (37,383)^2 = 1397,79$, то можно получить объективную оценку различия между ними.

Сравнение по критерию Фишера (см. § 4.01) при величине большей дисперсии $\sigma_1^2 = 2009,55$ и меньшей дисперсии $\sigma_2^2 = 1397,79$ показывает несущественное различие между ними:

$$F = 2009,55/1397,79 = 1,44,$$

что близко к единице при $\nu(1) = \infty$ и $\nu(2) = \infty$ на доверительном уровне $P_3 = 99,9\%$ (табл. 9П), который рекомендуется для оценок в подобных случаях. Если объем совокупности больше 120, как это

Таблица 1.42. Определение средней и сигмы распределения величин коэффициента вариации

	f	$p\%$	xf	$x - M$	$(x - M)f$	$(x - M)^2f$
10	89	50,85714	890	-21,8286	-1942,7454	42 407,4122
30	45	25,71429	1350	-1,8286	-82,2870	150,4700
50	19	10,85714	950	18,1714	345,2566	6 273,7958
70	9	5,14286	630	38,1714	343,5426	13 113,5020
90	3	1,71428	270	58,1714	174,5142	10 151,7353
110	1	0,57143	110	78,1714	78,1714	6 109,7678
130	3	1,71428	390	98,1714	294,5142	28 912,8713
150	3	1,71429	450	118,1714	354,5142	41 893,4393
170	2	1,14286	340	138,1714	276,3428	38 182,6716
190	1	0,57143	190	158,1714	158,1714	25 018,1918
	175	100,00000	5570		-0,0050	212222,7571

часто бывает, то оценку различия квадратов средней и сигмы можно делать без таблиц; для этого достаточно сравнить полученное значение F с числом 1,56. При $F \geq 1,56$ различие двух показателей существенно, и данное распределение не подходит для аппроксимации его показательным. Если $F < 1,56$, то при подходящей форме кривой на графике некоторое эмпирическое распределение может быть интерпретировано как показательное.

Рассмотрим также применение показательного распределения к важному для биологических исследований распределению величин коэффициента вариации 175 различных признаков растений. Каждая величина коэффициента вариации в свою очередь была получена путем обработки взвешенного ряда достаточного объема, и общее число первичных наблюдений, использованное для получения рассматриваемого ряда составляет более 70 300. В табл. 1.42 приведены: x — середина класса (или варианта) — величины коэффициента вариации с интервалом $s = 20$; f — частота; $p\%$ — частости эмпирического распределения; xf — произведения вариант на частоты; $x - M$, $(x - M)f$, $(x - M)^2f$ — отклонения вариант от средней, произведения отклонений, а также их квадратов с частотами. На основании данных табл. 1.42, которая служит также примером для способа «д» из § 1.10, вычисляем среднюю арифметическую: $M = \sum xf / N = 5570 / 175 = 31,828573$. Проверка: $\sum xf = MN$; $5570 = 31,828573 \cdot 175 = 5570,000275$. Разница, равная 0,000275, в данном случае может считаться несущественной. Правильность вычисления отклонений проверяем по равенству: $\sum (x - M)f = 0$; в нашем примере эта сумма равна -0,0050, что можно считать несущественной ошибкой.

Дисперсия:

$$\sigma^2 = \frac{\sum (x - M)^2f}{N - 1} = \frac{212222,7571}{175 - 1} = 1219,671.$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1219,671} = 34,92379.$$

Проверяем вычисление сигмы посредством факториальных моментов по способу полного суммирования, который более, чем другие способы из § 1.18, подходит для подобных асимметричных рядов. В табл. 1.43 подсчитаны суммы, необходимые для получения четырех факториальных моментов.

Дисперсия:

$$\sigma^2 = \frac{c^2}{N-1} \left(2s_2 - s_1 \left(1 + \frac{s_1}{N} \right) \right) = \frac{20^2}{175-1} \left(2 \cdot 831 - 366 \left(1 + \frac{366}{175} \right) \right) = 1219,625212; \quad \sigma = 34,92.$$

Первый условный момент: $m_1 = s_1/N = 366/175 = 2,0914287$.

Средняя арифметическая: $M = x_1 + c(m_1 - 1) = 10 + 20 \cdot (2,0914287 - 1) = 31,82857$, здесь x_1 — первая варианта из столбца 1 табл. 1.42.

Проверяем равенство средней и сигмы как условие показательного распределения:

$$F = \frac{\sigma^2}{M^2} = \frac{1219,671}{(31,828572)^2} = 1,2 < 1,56.$$

Из сравнения величины критерия Фишера с числом 1,56 заключаем, что ряд распределения величин коэффициента вариации можно представить показательным распределением.

Основной параметр (1.166):

$$\lambda = \frac{1}{31,8286} = 0,031418, \dagger$$

дальнейшие действия представлены в табл. 1.44, где обозначения те же, что и в табл. 1.41. Совпадение теоретических (столбец 6 табл.

Таблица 1.43. К вычислению факториальных моментов распределения величин коэффициента вариации

f	f_1	f_2	f_3	f_4
89	175	366	831	1962
45	86	191	465	1131
19	41	105	274	666
9	22	64	169	392
3	13	42	105	223
1	10	29	63	118
3	9	19	34	55
3	6	10	15	21
2	3	4	5	6
1	1	1	1	1
$N = 175$	$s_1 = 366$	$s_2 = 831$	$s_3 = 1962$	$s_4 = 4575$

Таблица 1.44. Частоты показательного распределения величин коэффициента вариации

Верхние границы класса, g	Середина класса, x	λg	$e^{-\lambda g}$	p'	$f' = Np$	$p'\%$
0	—	0	1	—	—	—
20	10	0,628	0,5337	0,4663	81,60	46,6
40	30	1,257	0,2845	0,2492	43,61	24,9
60	50	1,885	0,1518	0,1327	23,22	13,3
80	70	2,513	0,08103	0,07077	12,38	7,08
100	90	3,142	0,04320	0,03783	6,62	3,78
120	110	3,770	0,02305	0,02015	3,53	2,02
140	130	4,398	0,01229	0,01076	1,88	1,08
160	150	5,027	0,006558	0,005732	1,0	0,573
180	170	5,655	0,0035	0,003058	0,535	0,306
200	190	6,284	0,001866	0,001634	0,286	0,163
220	210	6,912	0,000996	0,00087	0,152	0,087
$c = 20$ $k = 11$					174,813	99,889

1.44) и эмпирических (столбец 1 табл. 1.43) частот удовлетворительное, что видно из величины критерия Колмогорова $K(\lambda) = 0,66529$, которая меньше первого критического уровня: 1,36.

§ 1.35. Распределение Парето

Ряды с односторонней кривой распределения и в первую очередь с резким падением частоты вариантов в начале кривой рекомендуется аппроксимировать распределением Парето, которое относится к непрерывному типу, определяется всего одним параметром и сравнительно просто по технике вычислений.

Например, число побегов на одной особи гелениума после деления чаще всего небольшое — один побег имели большинство, т. е. 310 особей (табл. 1.45), а два побега — только 13 особей. В этом месте происходит резкое снижение частоты с 310 до 13 и кривая образует почти прямой угол, закругленный на вершине (рис. 12), что типично для кривой распределения Парето, напоминающей сильно вогнутую гиперболу. Средняя арифметическая распределения побегов гелениума $M = 1,1947$, начальная, или минимальная, варианта ряда $x_1 = 1$ побегу и среднее квадратическое отклонение $\sigma = 0,77906$.

Последовательность расчета показателей и частот этого типа распределения следующая.

1. Для вычислений требуются средняя арифметическая, начальное значение ряда, частоты, середины и границы классов. Единственный параметр распределения находим по формуле

$$\alpha = \frac{M}{M - x_1}, \quad (1.171)$$

Таблица 1.45. Частоты распределения Парето

Число побегов на одно- летней особи гелениума		Число особей, f	$\ln x$	$\alpha \ln x$	$x^{-\alpha}$	$x_{i-1}^{-\alpha} - x_i^{-\alpha} = B$	$102,92 \cdot B = f'$	$f'\%$
Середина интервала	Границы интервала, x							
	0,5		-0,69316	1,1921	3,294			
1	1,5	310	0,40546	-0,69731	0,49795	2,796	287,76	84,9
2	2,5	43	0,9163	-1,5759	0,20682	0,29113	29,963	8,9
3	3,5	7	1,2528	-2,4546	0,11595	0,09087	9,3523	2,7
4	4,5	2	1,5041	-2,5868	0,075262	0,04069	4,1878	1,2
5	5,5	3	1,7048	-2,9319	0,053293	0,021969	2,2611	0,67
6	6,5	3	1,8718	-3,2191	0,04	0,013296	1,3684	0,4
7	7,5	1	2,015	-3,4654	0,031258	0,008739	0,89942	0,27
8	8,5					0,006045	0,62215	0,18
>8	∞						2,59	0,76
$k = 7$	$c = 1$	339					339,0	99,98

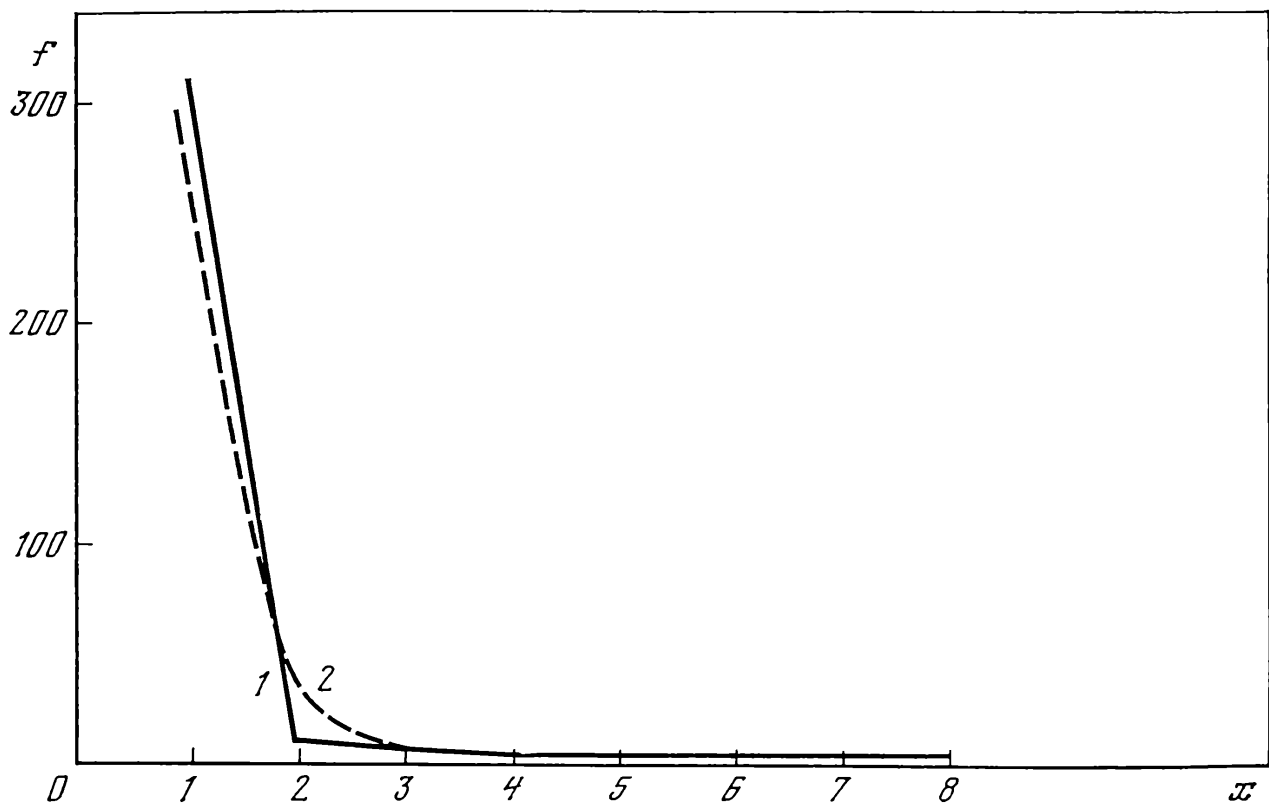


Рис. 12. Распределение однолетних особей гелениума осеннего по числу побегов на них (x)

f — частоты: 1 — эмпирические; 2 — кривой Парето;

где α — параметр распределения Парето; M — средняя арифметическая взвешенного эмпирического ряда; x_1 — начальное значение ряда. В данном случае вычисления будут вестись не по средним значениям классов, а по их границам. Левая наименьшая граница первого класса — 0,5, это число и принимаем за x_1 .

По формуле (1.171):

$$\alpha = \frac{1,1947}{1,1947 - 0,5} = 1,7198.$$

2. Медиана ряда равна:

$$M_e = 2^{\frac{1}{\alpha}} x_1 = x_1 \exp\left(\frac{\ln 2}{\alpha}\right); \quad (1.172)$$

$$M_e = 0,5 \cdot \exp\left(\frac{0,69315}{1,7198}\right) = 0,7482.$$

3. Частоты находим по формуле:

$$f' = Nx_1^\alpha (x_{i-1}^{-\alpha} - x_i^{-\alpha}), \quad (1.173)$$

где f' — частоты распределения Парето; x_1 — начальное значение ряда, в нашем примере равно 0,5; α — параметр распределения, равный в данном примере 1,7198; x_{i-1} — предыдущая граница класса; x_i — последующая граница класса.

Множитель перед скобками в этой формуле посредством логарифмов

рифмирования] вычислим сразу:

$$Nx_1^\alpha = 339 \cdot 0,5^{1,7198} = 102,92.]$$

4. Дальнейшие вычисления приведены в табл. 1.45. В столбце 4 натуральные логарифмы границ классов, в столбце 5 они умножены на $-\alpha$, в столбце 6 — экспоненты, или антилогарифмы, чисел 5 столбца. В столбце 7 записаны разности предыдущего и последующего чисел из столбца 6. Умножив их на 102,92, получаем в столбце 8 частоты распределения Парето, для наглядности представленные также в процентах (столбец 9). Более точная аппроксимация данного распределения приведена в § 1.30 по второму способу негативного биномиального распределения, однако для этого там потребовались более сложные расчеты.

Вычисление частот далее седьмого класса в данном случае бесполезно, так как частота не может быть менее 1, т. е. одной особи. Кривая Парето асимптотически приближается к оси x , и число вариантов может быть неограниченно большим. Класс > 8 побегов в таблице означает, что теоретически в совокупности может встретиться около 0,76% особей, на которых может быть от 9 до 16 побегов. Последнее число было найдено продолжением вычислений до 16 класса, после которого сумма столбца 8 достаточно точно приблизилась к $N = 339$.

Дисперсия, среднее квадратическое отклонение (σ) и коэффициент вариации (v) распределения Парето имеют смысл при условии, что $\alpha > 2$ или $2x_1 > M$ и вычисляются по формулам:

$$\sigma^2 = \frac{\alpha x_1^2}{(\alpha - 2)(\alpha - 1)^2}, \quad \sigma = \sqrt{\sigma^2}, \quad (1.174)$$

$$v = \frac{100}{\sqrt{\alpha(\alpha - 2)}}, \quad \text{или} \quad v = \frac{100(M - x_1)}{\sqrt{M(2x_1 - M)}}, \quad (1.175)$$

где обозначения те же, что и к формуле (1.171). В нашем примере $\alpha < 2$ и $2x_1$ меньше, чем величина средней арифметической, поэтому показатели по формулам (1.174), (1.175) не вычисляем.

§ 1.36. Равномерное распределение

Распределение дискретных, или непрерывных, величин с равной для всех из них вероятностью называется равномерным. Кривая распределения частот этого типа представляет собой прямую, параллельную оси абсцисс, ограниченную слева и справа значениями минимальной и максимальной вариант. Наиболее частым случаем применения этого типа распределения можно считать проверку на равную вероятность появления любой варианты из данной совокупности. Рассмотрим два практических примера.

В тиражах выигрышей спортивной лотереи «Спортлото 5 из 36» в 1977—1979 гг. числа от 1 до 36 появлялись с различной частотой (столбцы 1, 2 табл. 1.46). Определим, существенна ли разница

Таблица 1.46. Проверка соответствия частот чисел лотерей равномерному распределению

x	f	f^2	n	n'	$\frac{n}{N}$	$\frac{n'}{N}$	$d \cdot 1000$
1	41	1681	41	43,611	0,026	0,028	2
2	43	1849	84	87,2	0,053	0,055	2
3	34	1156	118	130,8	0,075	0,083	8
4	35	1225	153	174,4	0,097	0,111	14
5	46	2116	199	218,1	0,127	0,139	12
6	54	2916	253	261,7	0,161	0,167	6
7	58	3364	311	305,3	0,198	0,194	4
8	45	2025	356	348,9	0,227	0,222	5
9	42	1764	398	392,5	0,253	0,250	3
10	48	2304	446	436,1	0,284	0,278	6
11	53	2809	499	479,7	0,318	0,306	6
12	39	1521	538	523,3	0,342	0,333	9
13	43	1849	581	566,9	0,370	0,361	9
14	44	1936	625	610,5	0,398	0,388	10
15	56	3136	681	654,2	0,434	0,416	18
16	46	2116	727	697,8	0,463	0,444	19
17	42	1764	769	741,4	0,490	0,472	18
18	45	2025	814	785,0	0,518	0,500	18
19	51	2601	865	828,6	0,551	0,528	23
20	31	961	896	872,2	0,571	0,555	16
21	39	1521	935	915,8	0,595	0,583	12
22	37	1369	972	959,4	0,619	0,611	8
23	45	2025	1017	1003,1	0,648	0,639	9
24	51	2601	1068	1046,7	0,680	0,667	13
25	47	2209	1115	1090,3	0,710	0,694	16
26	39	1521	1154	1133,9	0,735	0,722	13
27	45	2025	1199	1177,5	0,764	0,750	14
28	43	1849	1242	1221,1	0,791	0,778	13
29	45	2025	1287	1264,7	0,820	0,806	14
30	36	1296	1323	1308,3	0,843	0,833	10
31	37	1369	1360	1351,9	0,866	0,861	5
32	49	2401	1409	1395,6	0,897	0,889	8
33	31	961	1440	1439,2	0,917	0,917	0
34	35	1225	1475	1482,8	0,939	0,944	5
35	50	2500	1525	1526,4	0,971	0,972	1
36	45	2025	1570	1570,0	1,000	1,000	0

$\Sigma f = 1570 \quad 70\,040$

между частотами появления этих чисел. Вычисления проводим в следующем порядке.

1. Находим среднюю частоту появления чисел

$$f' = \frac{\Sigma f}{k} = \frac{1570}{36} = 43,611112, \quad (1.176)$$

где f' — теоретическая частота равномерного распределения; Σf — сумма эмпирических частот; k — число классов распределения.

2. При помощи критерия хи-квадрат (более подробно об этом критерии см. главу 4) оценим степень соответствия эмпирических

частот равномерному распределению. Для этого эмпирические частоты возведем в квадрат (столбец 3) и получим сумму $\Sigma f^2 = 70040$. Величина критерия хи-квадрат равна:

$$\chi^2 = \frac{\Sigma f^2}{f'} - N = \frac{70040}{43,6111} - 1570 = 36, \quad (1.177)$$

где χ^2 — величина критерия хи-квадрат; Σf^2 — сумма квадратов эмпирических частот; f' — теоретическая частота равномерного распределения; $N = \Sigma f$ — объем совокупности, или сумма эмпирических частот.

Оценку найденной величины хи-квадрат делаем с помощью табл. 10П. Число степеней свободы равно: $\nu = k - 1 = 36 - 1 = 35$. При трех стандартных уровнях значимости: 0,05; 0,01; 0,001 табличные значения критерия равны соответственно 49,8; 57,3; 66,6, что больше найденной выше величины, равной 36. Следовательно, различия между частотами появления чисел в тиражах выигрышей случайны, и они появляются практически с равной вероятностью, иными словами их распределение подчиняется закону равномерного распределения.

3. Для контроля проверку соответствия эмпирических частот равномерному распределению сделаем также по критерию Колмогорова (см. о нем также в главе 4). В столбце 4 таблицы записаны накопленные эмпирические частоты, которые получены так: $41 + 43 = 84$; $84 + 34 = 118$; $118 + 35 = 153$ и т. д. В столбце 5 записаны накопленные теоретические частоты, которые начинаются со средней частоты $f' = 43,611$, к ней последовательно прибавляется ее же величина: $43,611 + 43,611 = 87,2$; $87,2 + 43,611 = 130,8$ и т. д. В столбцах 6, 7 частные от деления этих частот на объем совокупности $N = 1570$. В столбце 8 — разности чисел из предыдущих двух столбцов, для краткости написанные умноженные на 1000, $d_{max} = 23$ соответствует 0,023. Эту максимальную разность подставляем в формулу

$$\lambda = d_{max} \sqrt{N} = 0,023 \sqrt{1570} = 0,911. \quad (1.178)$$

В соответствии с полученной величиной лямбды, равной 0,91, по табл. 28П находим значение $1 - K(\lambda) = 0,379$. Если эта табличная величина будет больше критической вероятности 0,05, то данное распределение не отклоняется от теоретического, а если $1 - K(\lambda) < 0,05$, то исследуемое эмпирическое распределение существенно отклоняется от теоретического распределения частот. В нашем случае $1 - K(\lambda) = 0,379 > 0,05$. Следовательно, распределение частот чисел лотереи практически не отличается от равномерного, что подтверждает результат уже сделанной проверки при помощи критерия хи-квадрат. И поэтому все попытки некоторых любителей этой игры найти «счастливые» комбинации чисел являются заведомо и полностью бесплодными.

Рассмотрим второй пример применения равномерного распределения, относящийся к биологическим исследованиям. При помощи равномерного распределения можно также получать ответы на воп-

Таблица 1.47. Распределение 270 двулетних растений 22 сортов гелениума осеннего по высоте

Высота, см, x	Число особей, f	f^2	n	n'	n/N	n'/N	d
86	3	9	3	30	0,011	0,111	0,100
98	10	100	13	60	0,048	0,222	0,174
110	21	441	34	90	0,126	0,333	0,207
122	19	361	53	120	0,196	0,444	<u>0,248</u>
134	41	1681	94	150	0,348	0,555	0,207
146	92	8464	186	180	0,689	0,667	0,022
158	57	3249	243	210	0,900	0,778	0,122
170	23	529	266	240	0,985	0,889	0,096
182	4	16	270	270	1,000	1,000	0
	270	14 850					

росы следующего типа: правомерно ли считать сортовым признаком, например, высоту растений у совокупности сортов какого-либо культивируемого вида. В табл. 1.47 приведено распределение по 9 классам высоты 270 двулетних особей 22 сортов гелениума осеннего. Требуется выяснить, различаются ли достоверно растения по высоте в данной совокупности и, следовательно, может ли служить высота растений сортовым признаком. Средняя частота распределения равна (1.176): $f' = 270/9 = 30$. Сравнение по критерию хи-квадрат (1.177): $\chi^2 = 14850/30 - 270 = 225$ (по табл. 10П при $\nu = 9 - 1 = 8$ и уровне значимости 0,001 критическое значение 26,1) и по критерию Колмогорова: $\lambda = 0,248\sqrt{270} = 4,075$ (что соответствует нулевой вероятности $0 < 0,05$) показало, что данное распределение частот существенно отличается от равномерного. Следовательно, высота растений у данной совокупности сортов может служить сортовым признаком.

§ 1.37. Логнормальное распределение

Логнормальное распределение называют также логарифмически нормальным, или типом S_L из семейства распределений Джонсона. Весьма пластичное логнормальное распределение часто применяется для аппроксимации распределений многих биологических признаков, особенно тех из них, которые обладают умеренной положительной асимметрией.

Из различных способов вычисления частот этого распределения рассмотрим наиболее рациональный — способ процентилей на примере распределения величин коэффициента вариации 165 биологических признаков. В § 1.34 этот резко асимметричный ряд был с удовлетворительной точностью аппроксимирован показательным распределением. Однако слишком большой шаг интервала в этом ряду: $c = 20$ заставлял предполагать, что в начале ряда, где осо-

Таблица 1.48. К расчету частот логнормального распределения

Нижняя граница класса, g	Частота f	Накопленная частота $f\Sigma$	$g-\varepsilon$	$\ln(g-\varepsilon)$	$\eta \cdot (5)$	$z=\gamma+(6)$	P	$P_{i+1}-P_i$	$f'=N \cdot (9)$	$f, \%$
1	2	3	4	5	6	7	8	9	10	11
0	12	12	4,15	1,423	2,127	-2,56	0,005234	0,0801	13,22	8,0
5	26	38	9,15	2,214	3,310	-1,37	0,08534	0,1504	24,82	15,0
10	27	65	14,15	2,649	3,960	-0,72	0,23576	0,1578	26,04	15,8
15	24	89	19,15	2,952	4,413	-0,27	0,39358	0,1383	22,82	13,8
20	14	103	24,15	3,184	4,760	0,08	0,53188	0,1087	17,94	10,9
25	14	117	29,15	3,373	5,043	0,36	0,64058	0,0852	14,06	8,5
30	10	127	34,15	3,531	5,279	0,60	0,72575	0,0624	10,30	6,2
35	7	136	39,15	3,667	5,482	0,80	0,78814	0,0483	7,97	4,8
40	4	138	44,15	3,788	5,663	0,98	0,83646	0,0364	6,01	3,6
45	10	148	49,15	3,895	5,823	1,14	0,87286	0,0286	4,72	2,9
50	3	151	54,15	3,992	5,968	1,29	0,90147	0,0207	3,42	2,1
55	2	153	59,15	4,080	6,100	1,42	0,92220	0,0160	2,64	1,6
60	5	158	64,15	4,161	6,221	1,54	0,93822	0,0123	2,03	1,2
65	1	159	69,15	4,236	6,333	1,65	0,95053	0,0103	1,70	1,0
70	1	160	74,15	4,306	6,437	1,76	0,96080	0,0070	1,16	0,7
75	2	162	79,15	4,371	6,535	1,85	0,96784	0,0066	1,09	0,7
80	1	163	84,15	4,433	6,627	1,95	0,97441	0,0044	0,73	0,4
85	1	164	89,15	4,490	6,713	2,03	0,97882	0,0038	0,63	0,4
90	1	165	94,15	4,545	6,795	2,11	0,98257	0,0032	0,53	0,3
95	0	165	99,15	4,597	6,872	2,19	0,98574	0,0024	0,40	0,2
100			104,15	4,646	6,946	2,26	0,98809			

$c = 5$

165

$k = 20$

162, 23

98,1

бенно велика плотность вариантов, могут быть скрыты существенные детали кривой данного распределения. По тем же данным был заново получен взвешенный вариационный ряд с шагом интервала $c = 5$, асимметрия которого оказалась значительно меньшей и эмпирическая кривая стала двускатной, что больше соответствует реальным фактам. При этом пришлось отбросить те варианты, которые были больше 100%. Объем ряда поэтому стал $N = 165$ при числе классов $k = 20$ (табл. 1.48). Расчеты частот логнормального распределения выполнены в следующем порядке.

1. Составление ряда накопленных частот: $12 + 26 = 38$; $38 + 27 = 65$; $65 + 24 = 89$ и т. д. до последней частоты, которая должна быть равна объему ряда (столбец 3).

2. Вычисление номеров процентильных вариантов:

$$i_P = P(N + 1), \quad (1.179)$$

где i_P — номер процентильной варианты; P — величина процентиля; N — объем ряда. За величины процентилей приняты медиана $P = 0,5$ и симметричные ей $P = 0,05$, $P = 0,95$. По формуле (1.179): $i_{0,05} = 0,05(165 + 1) = 8,3$; $i_{0,5} = 0,5(165 + 1) = 83$; $i_{0,95} = 0,95(165 + 1) = 157,7$.

3. Значения процентильных вариантов:

$$x_P = (i_P - f_H) \frac{c}{f} + g, \quad (1.180)$$

где x_P — значение процентильной варианты; i_P — номер процентильной варианты; f_H — накопленная частота класса, предшествующего процентильному; c — классовый интервал; f — частота процентильного класса (не накопленная); g — левая, или меньшая, граница процентильного класса; $x_{0,05} = (8,3 - 0) \frac{5}{12} + 0 = 3,46$; $x_{0,5} = (83 - 65) \frac{5}{24} + 15 = 18,75$; $x_{0,95} = (157,7 - 153) \frac{5}{5} + 60 = 64,7$.

Таким образом, процентильные варианты находятся в 1, 4 и 13 классах ряда (сверху вниз, по строкам столбца 1 табл. 1.48).

4. Параметры логнормального распределения:

$$\eta = 1,645 \left[\ln \left(\frac{x_{0,95} - x_{0,5}}{x_{0,5} - x_{0,05}} \right) \right]^{-1} \quad (1.181)$$

$$\gamma = \eta \ln \left(\frac{1 - e^{-1,645/\eta}}{x_{0,5} - x_{0,05}} \right), \quad (1.182)$$

$$\varepsilon = x_{0,5} - e^{-\gamma/\eta}, \quad (1.183)$$

$$\eta = 1,645 \left[\ln \left(\frac{64,7 - 18,75}{18,75 - 3,46} \right) \right]^{-1} = 1,495;$$

$$\gamma = 1,495 \ln \left(\frac{1 - e^{-1,645/1,495}}{18,75 - 3,46} \right) = -4,6820;$$

$$\varepsilon = 18,75 - e^{4,682/1,495} = -4,15.$$

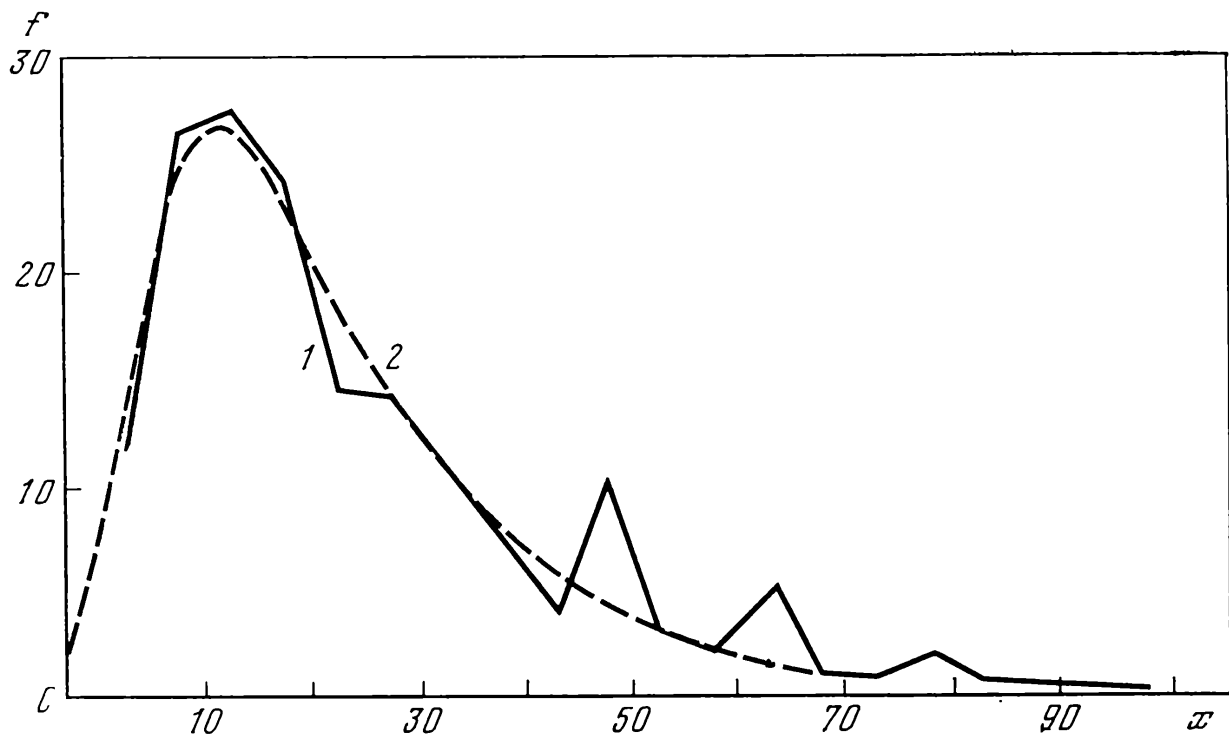


Рис. 13. Логнормальное распределение величин коэффициента вариации (x) биологических признаков

f — частоты: 1 — эмпирические, 2 — теоретические

5. Нормированные отклонения (z) левых (нижних, или меньших) границ классов (g), вычисленные по формуле:

$$z = \gamma + \eta \ln (g - \epsilon), \quad (1.184)$$

записаны в столбце 7.

6. Из табл. 27П в соответствии с величинами z в столбец 8 табл. 1.48 выписываем значения функции Φ , т. е. площади под интегральной кривой нормального распределения. Вычитая затем предыдущее значение из последующего, в столбце 9 получим частоты логнормального распределения, выраженные в долях единицы.

7. Произведения чисел столбца 9 на объем ряда $N = 165$ (столбец 10) представляют собой частоты логнормального распределения, которые также выражены в процентах (столбец 11). Совпадение эмпирических и теоретических частот достаточно близкое (рис. 13), что показывает и величина критерия $K(\lambda) = 0,40336$.

На этом задача аппроксимации данного распределения может быть закончена, но ввиду важности рассматриваемого ряда для оценки вариабельности признаков растений в целом вычислим его другие параметры.

1. Средняя арифметическая по серединам классов (1.13):

$$M = 4057,5/165 = 24,5909.$$

2. Среднее квадратическое отклонение (1.24а):

$$\sigma = \sqrt{\frac{60453,637}{165 - 1}} = \sqrt{368,6197} = 19,8958.$$

3. Проверка вычисления средней и сигмы по формулам (1.70), (1.69a):

$$M = 2,5 + 5 (5,4182 - 1) = 24,591;$$

$$\sigma^2 = \frac{25}{164} \left(2 \cdot 4078 - 894 \left(1 + \frac{894}{165} \right) \right) = 368,6196; \quad \sigma = 19,896.$$

4. Моменты условные (1.68):

$$m_1 = \frac{894}{165} = 5,4182; \quad m_2 = \frac{2 \cdot 4078}{165} - 5,4182 = 44,0121;$$

$$m_3 = 2 \left(3 \left(\frac{16852}{165} - 0,5 \cdot 44,0121 \right) - 5,4182 \right) = 469,9268;$$

$$m_4 = 6 \left(\frac{4 \cdot 63841}{165} - 5,4182 - 469,9268 \right) - 11 \cdot 44,0121 = 5949,7593.$$

Центральные моменты (1.56) — (1.58):

$$\mu_2 = 44,0121 - 5,4182^2 = 14,6552;$$

$$\mu_3 = 469,9268 + 5,4182(2 \cdot 29,3569 - 3 \cdot 44,0121) = 72,6508;$$

$$\mu_4 = 5949,7593 + 5,4182(3 \cdot 5,4182(2 \cdot 44,0121 - 29,3569) - 4 \cdot 469,9268) = 931,9984.$$

5. Коэффициент вариации эмпирического ряда:

$$v = \frac{19,9}{24,6} \cdot 100 = 81 \%.$$

Величина: $\frac{M^2}{\sigma^2} = \frac{604,7}{368,6} = 1,76 > 1,56$ (см. § 1.34), поэтому данное эмпирическое распределение нельзя аппроксимировать показательным.

6. Уравнение, по которому можно вычислить частоту для любой величины x в пределах области аппроксимации логнормальным распределением:

$$f' = \frac{Nc\eta}{(x - \varepsilon) \sqrt{2\pi}} \exp \left[-0,5\eta^2 \left(\frac{\gamma}{\eta} + \ln(x - \varepsilon) \right)^2 \right], \quad (1.185)$$

где N — объем ряда; c — классовый интервал; x — случайная величина, любая варианта ряда, в том числе g — граница класса; η , γ , ε — параметры распределения.

В рассматриваемом примере вычислим частоту для $x = 33\%$, т. е. теоретическую частоту появления рядов распределений биологических признаков с коэффициентом вариации 33% , что соответствует частоте встречаемости рядов с нормальным распределением:

$$\begin{aligned} f' &= \frac{165 \cdot 5 \cdot 1,495}{(33 + 4,15) \sqrt{2\pi}} \exp \left(-0,5 \cdot 1,495^2 \left(\frac{-4,682}{1,495} + \ln(33 + 4,15) \right)^2 \right) = \\ &= 165 \cdot 0,08027 = 13,24. \end{aligned}$$

где: $1/\sqrt{2\pi} = 0,39894$. Таким образом, в данной совокупности может встретиться около 8% рядов, вариабельность которых близка к 33% .

Используя полученное значение среднего квадратического отклонения, построим интервал нормы variability признаков растений: $M \pm \sigma = 24,59 \pm 19,9$, т. е. от 4,69 до 44,49, далее: 44,49 — 64,39—84,29—104,19. Округленно шестибалльная шкала variability выглядит так: 0—4,69; 4,7—24,59; 24,6—44,49; 44,5—64,39; 64,4—84,29; 84,3—104,2 (величины коэффициента вариации в процентах).

Иногда для точного определения типа кривой из семейства распределений Джонсона пользуются следующими критериями:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \beta_2 = \frac{\mu_4}{\mu_2^2}. \quad (1.186)$$

Подставив в эти формулы найденные значения центральных моментов, получим:

$$\beta_1 = \frac{72,6508^2}{14,6552^3} = 1,6769; \quad \beta_2 = \frac{931,9984}{14,6552^2} = 4,3394.$$

Однако, как показывает наш опыт, аппроксимация логнормальным распределением бывает обычно достаточно точной, несмотря на то, что по критериям (1.186) ряд может относиться к другому типу кривой, как это имеет место в данном случае, в котором согласно критериям следовало бы применить тип кривой S_B . По данным аппроксимации через процентиля можно получить также приближенную оценку коэффициента вариации распределения S_L Джонсона по формуле:

$$v = 100 \sqrt{\exp(\eta^{-2}) - 1}, \quad (1.187)$$

где η вычисляется по (1.181). Для нашего примера:

$$v = 100 \sqrt{\exp(1,495^{-2}) - 1} = 75,1\%.$$

КОРРЕЛЯЦИЯ

Между величинами может существовать более точная, или функциональная, связь, когда одному значению аргумента (x) соответствует одно определенное значение функции (y), и менее точная — корреляционная связь, когда одному конкретному значению аргумента соответствует приближенное значение, или некоторое множество значений функции, в той или иной степени близких друг к другу. Близость в этом множестве значений функции друг к другу соответствует понятию силы, или тесноты корреляционной связи. Чем больше сила корреляционной связи, тем ближе эта связь к функциональной.

Объекты ботанических исследований (популяции, особи, органы, клетки) всегда в той или иной степени неоднородны по своим наследственным особенностям и, кроме того, в любом, даже тщательно поставленном эксперименте его объекты испытывают неучитываемые воздействия многих факторов внешней среды. Поэтому между признаками объектов ботанических исследований бывают исключительно корреляционные связи. Например, большей длине корня у высших растений, как правило, соответствует в общем большая высота стебля; в этом случае имеет место положительная, или прямая, корреляция. Чем растения крупнее, больше по размерам или по весу, тем меньшее их число может произрастать на одной и той же единице площади участка земли; в этом случае существует отрицательная, или обратная, корреляция. В одной и той же парной связи может быть и прямая, и обратная корреляция. Например, высота растений и число соцветий на них у гелениума осеннего находятся в прямой связи, но до некоторого предела, после которого при дальнейшем увеличении высоты растений происходит не увеличение, а наоборот, уменьшение числа соцветий на них. Поэтому при изучении корреляций следует брать достаточно широкие пределы изменений обоих признаков, чтобы была возможность выявить и такие двузначные связи.

Сложность изучения корреляционных связей заключается также и в том, что все признаки (например, у растения) в различной степени взаимосвязаны и при исследовании методом парной корреляции можно считать доказанной лишь ту связь между двумя признаками, механизм которой биологически понятен; в противном случае можно установить корреляцию там, где ее нет, ибо изменения двух признаков могут более сильно зависеть от изменения иного (третьего) признака или от совокупности некоторых других признаков. Поэтому более правильным было бы совместное, одновремен-

ное исследование связей всего комплекса признаков данного растения или явления с применением методов множественной корреляции и корреляционных плеяд.

Методы вычисления показателей корреляции располагаются в следующем порядке: ранговые показатели корреляции, наиболее простые по технике вычисления (§ 2.02, 2.03), обычный коэффициент корреляции (§ 2.04, 2.05), схемы расчетов показателей криволинейной связи (§ 2.06—2.12), некоторые частные случаи корреляционных связей (§ 2.13—2.16).

Для проверки гипотезы о взаимосвязи между двумя признаками, которые могут быть выражены количественно, вычисляют различные показатели силы и достоверности этой связи. Для невзвешенных рядов (см. 1.01) составление корреляционной решетки (пример ее см. табл. 2.01) не требуется, так как проще вычислить коэффициент корреляции без ее составления. Если объемы выборок значительны и требуют группировки вариант во взвешенные ряды, то для вычисления различных показателей связи — коэффициента корреляции, корреляционного отношения, полихорического показателя связи желательно, особенно при отсутствии вычислительной техники, составить корреляционную решетку. Вычисление эмпирической линии регрессии сильно облегчает возможность выявления формы связи, чтобы далее решать, какой из показателей связи следует применить в данном конкретном случае. Если связь прямолинейная, то вычисляют коэффициент корреляции, показатель корреляции Спирмена, коэффициент корреляции Кендэла, а при криволинейной связи вычисляют корреляционное отношение и полихорический показатель связи независимо от объема выборки. Точки эмпирической линии регрессии, вычисленные по способу, приведенному в § 2.01, используются также при вычислении корреляционного отношения. При вычислении прямого корреляционного отношения требуется получить средние значения нескольких y для одного определенного значения x , а при вычислении обратного корреляционного отношения вычисляют усредненные значения нескольких x для определенного значения y . Таким образом, существует два вида эмпирической линии регрессии: y/x и x/y , т. е. y по x и x по y . Вычисление точек эмпирической линии регрессии производится также и в том случае, когда корреляционное отношение находят для невзвешенного ряда.

При исследовании корреляционных связей между признаками любых, в том числе и биологических объектов, вначале желательно выявить тип распределения данных (см. главу 1). При этом чем ближе эмпирическое распределение к нормальному или хотя бы к симметричному, тем больше оснований для применения методов корреляционного анализа, и наоборот. Однако во многих случаях эту проверку нельзя выполнить из-за недостаточного объема исходных данных, что определяет в дальнейшем соответствующую надежность результатов.

Далее следует выявить степень криволинейности изучаемой связи, что можно сделать визуально по графику данной эмпирической

зависимости или аналитически (§ 2.07). Затем вычисляют избранные показатели связи и обязательно их ошибки для оценки достоверности найденных величин показателей. Недостоверные величины показателей корреляции из дальнейшего анализа в общем случае исключают.

Методы статистических исследований, изложенные в главе 1, могут быть отнесены к общему универсальному этапу математической обработки данных. Задача этого этапа — первичная обработка фактов, в результате которой ожидается получить основные характеристики массива данных: среднюю арифметическую, дисперсию, среднее квадратическое отклонение, ошибку средней, коэффициент вариации и показатель точности опыта. Перечисленные показатели позволяют судить об уровне и степени варьирования данных, оценить их полноту (репрезентативность), надежность и пригодность для дальнейших исследований и выводов. Методы корреляционного анализа, сравнений, регрессии и многие другие, изложенные в главах 2—5, относятся к специальному этапу статистических исследований, на котором, в отличие от общего, требуется выдвижение четко сформулированной рабочей гипотезы, которую надлежит доказать или опровергнуть при помощи биометрических методов. В частности, посредством вычисления показателей корреляции и оценки их достоверности определяют наличие или отсутствие связи между различными признаками объектов.

§ 2.01. Корреляционная решетка и эмпирическая линия регрессии

При составлении корреляционной решетки оба ряда данных надо сгруппировать в сопряженные между собой взвешенные ряды. Составление корреляционной решетки и вычисление точек эмпирической линии регрессии производится в следующем порядке.

1. Перед составлением сопряженных вариационных взвешенных рядов следует назначить аргумент (x) и функцию (y), т. е. определить, что является соответственно причиной и следствием в изучаемой взаимосвязи. Это поможет лучше ориентироваться в расчетах показателей связи, хотя само по себе является приемом довольно условным, так как в биологии часто исследуется связи, зависящие не друг от друга, а от чего-то третьего. Многие из парных взаимосвязей позволяют легко установить, какой член связи у них является аргументом. Но в некоторых случаях правильность назначения аргумента определяется лишь после вычислений и заключается в том, что у рассматриваемых двух рядов величина прямого корреляционного отношения будет больше, чем у обратного корреляционного отношения. Если обратное корреляционное отношение окажется больше прямого, за аргумент принимают другой из двух рядов. Вычислений при этом повторять нет необходимости, так как величины обоих корреляционных отношений не изменятся. При исследовании взаимосвязи: продолжительность вегетационного периода — вес 1000 семян у сортов ячменя (табл. 2.01) за аргумент (x) предположительно

принимается продолжительность вегетационного периода, которая может оказывать влияние на вес семян (y).

2. Вариационный ряд, который будет принят за функцию, лучше располагать в вертикальном ряду корреляционной решетки, числа ряда при этом возрастают снизу вверх. Значения аргумента располагаются в горизонтальном ряду, они возрастают слева направо. Такое же расположение рядов принимается при их графическом представлении. В табл. 2.01 приводятся границы классов и их срединные значения, т. е. варианты x , y , и затем частоты. Исходные данные разносятся по классам корреляционной решетки так же, как и при составлении отдельного вариационного ряда (§ 1.01) с той разницей, что здесь каждая точка в клетке (частота) соответствует определенным значениям обоих рядов одновременно. Сумма частот столбцов должна равняться сумме частот строк и объему выборки, в данном случае $N = 214$.

3. Расчет эмпирической линии регрессии y/x , которая используется при вычислении прямого корреляционного отношения, ведется путем определения взвешенных средних арифметических для отдельных участков ряда y или групп вариантов, имеющих частоты. Точка 37,0 (табл. 2.01) получена так:

$$\frac{40 \cdot 1 + 34 \cdot 1}{2} = 37,0.$$

4. Эмпирическая линия регрессии x/y (рис. 14) вычисляется в том случае, если требуется вычислить обратное корреляционное отношение, т. е., когда правильность первоначального назначения аргумента вызывает сомнение. Точки x/y получены путем определения групповых средних в горизонтальных рядах корреляционной решетки. Например,

$$\frac{65 \cdot 1 + 70 \cdot 2 + 75 \cdot 3 + 80 \cdot 7 + 85 \cdot 1}{14} = 76,8;$$

$$\frac{65 \cdot 5 + 70 \cdot 8 + 75 \cdot 10 + 80 \cdot 11 + 85 \cdot 2 + 90 \cdot 2}{38} = 75,4 \text{ и т. д.}$$

5. Вычисленные точки эмпирической линии регрессии y/x наносим на график (рис. 14) и соединяем линией. Полученная эмпирическая линия позволяет заключить, что связь между изучаемыми признаками положительная, так как при возрастании аргумента функция возрастает и близка к прямолинейной. Следовательно, в данном случае для определения степени силы связи можно применить коэффициент корреляции (см. § 2.05).

При обработке невзвешенных малочисленных рядов корреляционная решетка не составляется, однако график зависимости y от x по непосредственным данным вычерчивается для того, чтобы можно было судить о форме данной взаимосвязи, руководствуясь рис. 15. Вычерчивание такого графика перед вычислением показателя корреляции часто позволяет легко найти правильный метод обработки материала. Многие корреляционные связи, не слишком отклоняющиеся от прямолинейных, могут быть удовлетворительно

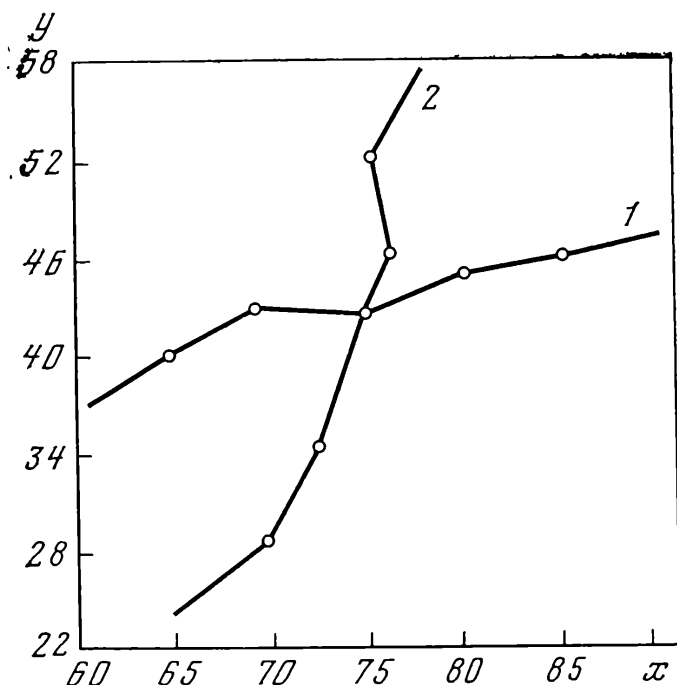
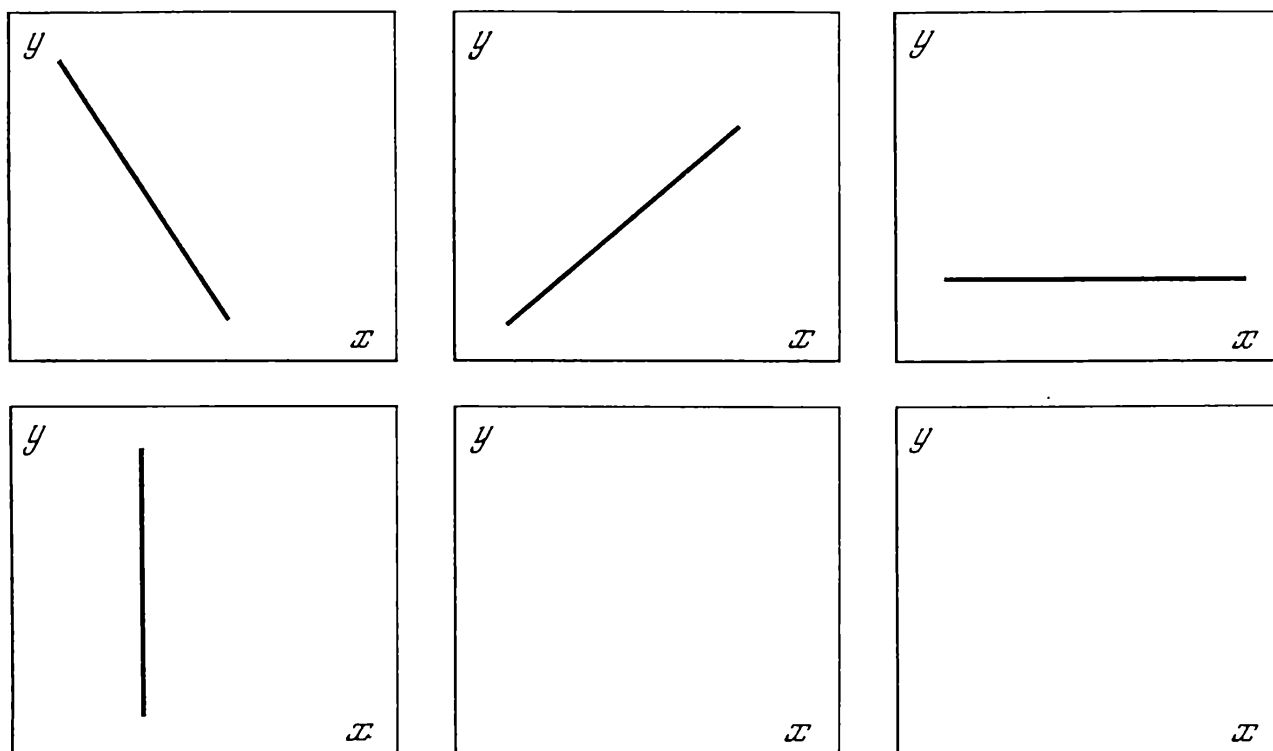


Рис. 14. Эмпирические линии регрессии

1 — вес 1000 семян (y) в зависимости от вегетационного периода: y/x
2 — продолжительность вегетационного периода (x) в зависимости от веса семян; x/y

Рис. 15. Различные формы корреляционных связей



характеризованы коэффициентом корреляции, который вычисляется проще, чем корреляционное отношение. Допустимая степень отклонения корреляционной связи от прямолинейной в таких случаях определяется при помощи критериев криволинейности (см. § 2.07).

§ 2.02. Показатель корреляции рангов по Спирмэну

Сила корреляционной связи между двумя признаками при помощи рассматриваемого показателя измеряется в основном для небольших невзвешенных рядов, которые представляют собой качественные оценки признаков.

Численность рядов в принципе может быть любой, однако ранжирование выборки большого объема становится затруднительным. Достаточную точность вычисления обеспечивает формула:

$$\rho = 1 - \frac{6(\sum \delta^2 + B_x + B_y)}{N(N^2 - 1)}, \quad (2.01)$$

где ρ — показатель корреляции рангов; N — число пар вариантов корреляционных рядов; B_x — поправка на объединение рангов в ряду x по формуле (2.03); B_y — поправка на объединение рангов в ряду y по формуле (2.03); $\sum \delta^2$ — сумма квадратов попарных разностей рангов.

Если в рядах x и y отсутствуют объединенные ранги, показатель корреляции рангов рассчитывается по формуле:

$$\rho = 1 - \frac{6\sum \delta^2}{N(N^2 - 1)}, \quad (2.02)$$

где обозначения те же, что и к формуле (2.01). В биометрических исследованиях в подавляющем большинстве случаев применяют формулу (2.02).

На следующем примере показан один из возможных случаев применения показателя корреляции рангов. Требуется объективно подобрать из числа опытных специалистов группу экспертов для оценки сортов декоративных растений. Для этого всем этим специалистам предлагается оценить 13 сортов одной цветочной культуры (вообще желательно для этой цели брать бóльшую выборку: 20—30 сортов или других образцов). Оценка ведется по декоративно-хозяйственным признакам растений и выражается в сумме баллов. Затем между оценками всех специалистов находят попарно показатель корреляции рангов и отбирают в качестве экспертов тех лиц, оценки которых в наибольшей степени совпадают. При этом требуется вычислить показатели корреляции, число которых равно: $N(N - 1)/2$, где N — число кандидатов на должность эксперта. Для примера вычислим показатель корреляции рангов между оценками только двух экспертов: x и y (столбцы 1 и 2 табл. 2.02).

1. Ранжируем данные рядов x и y от бóльшего к меньшему, присваивая вариантам порядковые номера. Варианта 72 ряда x получает ранг 1, варианта 67 — ранг 2, варианты 64 и 64 должны были бы получить ранги 3 и 4, но так как они имеют равное значение, то берется средний ранг для обеих вариантов: $(3 + 4)/2 = 3,5$. Остальные ранги рядов x и y приведены в столбцах 3, 4 табл. 2.02.

2. Получаем разности рангов без учета их знаков (столбец 5).

3. Возводим разности в квадрат и суммируем, получаем

$$\sum \delta^2 = 479 \text{ (столбец 6)}.$$

4. Вычисляем поправки по формуле:

$$B_x \text{ (или } B_y) = \sum \frac{n^3 - n}{12}, \quad (2.03)$$

где B_x или B_y — поправки на объединение рангов в ряду x или y ; n — число рангов в каждой группе объединенных рангов; \sum — знак

суммирования всех значений $(n^3 - n)/12$, вычисленных для отдельных групп.

Число групп объединенных рангов в ряду x равно 2. В каждой из групп по 2 ранга: 8,5 и 8,5; 3,5 и 3,5, отсюда по формуле (2.03):

$$B_x = \frac{(2^3 - 2) + (2^3 - 2)}{12} = 1.$$

В ряду y одна группа объединенных рангов с двумя рангами в ней: 3,5 и 3,5, отсюда по формуле (2.03):

$$B_y = \frac{2^3 - 2}{12} = 0,5.$$

5. Подставляем полученные значения в формулу (2.01):

$$\rho = 1 - \frac{6(479 + 1 + 0,5)}{13(13^2 - 1)} = -0,32.$$

Вычисляя по тем же данным, но без поправок на объединенные ранги по формуле (2.02), получим:

$$\rho = 1 - \frac{6 \cdot 479}{13(13^2 - 1)} = -0,32.$$

Таким образом, для оценки силы связи в данном случае можно было бы пользоваться и формулой (2.02).

Достоверность показателя корреляции рангов оценивается по формуле:

$$t = \rho \sqrt{\frac{N-2}{1-\rho^2}}, \quad (2.04)$$

где t — критерий достоверности Стьюдента; число степеней свободы для его оценки по табл. 3П определяется по выражению: $\nu = N - 2$; ρ — показатель корреляции рангов; N — объем ряда x .

Достоверность показателя корреляции рангов для приведенного примера по формуле (2.04):

$$t = 0,32 \sqrt{\frac{13-2}{1-0,32^2}} = 1,286.$$

По табл. 3П при $P'_1 = 0,95$ и $\nu = 13 - 2 = 11$, $t = 2,201$. Следовательно, полученный показатель корреляции рангов недостоверен и кандидаты на должность эксперта x и y оба не подходят или не подходит один из них, а кто именно, должны указать остальные показатели корреляции рангов между оценками других экспертов, рассмотренные совокупно.

Достоверность показателя корреляции рангов можно определить также по табл. 7П. При $N = 13$ вычисленное значение должно быть равно или больше табличного, которое равно 0,48. Следовательно, подтверждается сделанный выше вывод о недостоверности показателя корреляции. Его критические значения при объеме совокупности N можно также рассчитать по формулам: для 5% уровня значимости

$$\rho = \frac{1,96}{\sqrt{N-1}} \left(1 - \frac{0,16}{N-1} \right); \quad (2.05)$$

для 1% уровня значимости

$$\rho = \frac{2,58}{\sqrt{N-1}} \left(1 - \frac{0,69}{N-1} \right). \quad (2.06)$$

Последние формулы дают более строгую оценку. Если вычисленное значение ρ будет меньше, чем полученное по формуле (2.05) на 5% уровне значимости, то он недостоверен. Достоверным показатель корреляции рангов считается лишь тогда, когда он больше критического значения, полученного по формуле (2.06) на 1% уровне значимости.

§ 2.03. Коэффициент корреляции рангов по Кендэлу

Сила корреляционной связи между двумя невзвешенными рядами может быть определена посредством вычисления коэффициента корреляции рангов. Рекомендуется его вычислять для рядов не менее чем из 10 вариантов. Численность ряда в большую сторону теоретически может быть любой, однако ранжирование рядов с большим числом вариантов при ручном счете становится затруднительным, поэтому обычно применяют коэффициент Кендэла к небольшим выборкам. Вычисление производится по общей формуле:

$$\tau = B / \sqrt{\left[\frac{N(N-1)}{2} - B_x \right] \left[\frac{N(N-1)}{2} - B_y \right]} \quad (2.07)$$

где τ — коэффициент корреляции рангов по Кендэлу; B — сумма рангов, пример ее вычисления приводится в табл. 2.03;

N — объем выборки; B_x , B_y — поправки на объединение рангов, которые находим по формуле (2.08).

Если среди вариантов в обоих рядах нет совпадающих величин, следует обратиться к более простой формуле (2.12) вычисления коэффициента корреляции Кендэла.

Коэффициент корреляции рангов является более строгой оценкой связи, чем показатель корреляции рангов, так как по абсолютному значению $|\rho| > |\tau|$ для одних и тех же данных. На примере, приведенном в § 2.02, вычислим коэффициент корреляции рангов по формуле (2.07).

1. Рекомендуется расположить все значения ряда x , начиная с меньшего к большему; сопряженные значения y ранжируем вместе с x , т. е. передвигаем одновременно, они также займут соответствующее им положение.

2. Всем значениям обоих рядов от большого к меньшему присваиваем порядковые номера (ранги) так, как это описано в § 2.02, пункт 1. В примере предыдущего параграфа ряды уже ранжированы (столбцы 3, 4 табл. 2.02), остается лишь расположить ряды по порядку возрастания рангов ряда x (табл. 2.03).

3. Оценку рангов производим слева направо парами сопряженных рангов: в каждой паре один ранг из ряда x и другой, соответствующий ему, из ряда y . Каждая пара рангов сравнивается со всеми стоящими справа от нее. При сравнении могут возникнуть три

Таблица 2.03. Ранги и их оценка для вычисления коэффициента корреляции Кендэла

Ранги													
x	1	2	3,5	3,5	5	6	7	8,5	8,5	10	11	12	13
y	10	12	1	3,5	13	11	5	6	7	8	3,5	9	2
Оценка рангов													
Больше	+3	+1	+9	+7	0	0	+4	+2	+2	+1	+1	0	$\Sigma = +30$
Меньше	-9	-10	0	-1	-8	-7	-2	-2	-2	-2	-1	-1	$\Sigma = -45$
Сумма оценок: $B = +30 + (-45) = -15$.													

случая: а) ранг y левой из сравниваемых двух пар меньше соответствующего ранга y из правой пары, в этом случае записываем $+1$; б) ранг y из левой пары больше соответствующего ему ранга из правой пары, в этом случае записываем -1 ; в) ранги сравниваемых пар совпадают в одноименном ряду, в этом случае записываем 0. Начиная с пары рангов 1—10, сравнивая ее со всеми 12 парами, стоящими справа. В паре 2—12 ранг $y = 10$ из левой пары меньше ранга $y = 12$, записываем $+1$; в следующей паре: 3,5—1 левый ранг $y = 10$ больше правого ранга $y = 1$, записываем: -1 и т. д. Если ранги совпадают, например при сравнении восьмой пары с девятой: 8,5—6 и 8,5—7, или четвертой пары с одиннадцатой: 3,5—3,5 и 11—3,5, записываем 0. Затем подсчитываются суммы положительных и отрицательных единиц, что даст значение B . Указанный алгоритм поясняет принцип построения коэффициента и пригоден для программирования и вычисления на ЭВМ.

При ручном счете рекомендуется поступать следующим образом. Для оценки любой пары рангов, начиная с первой, подсчитываем, сколько рангов y из находящихся справа от нее имеют значения большие, чем значение ранга y , находящееся в оцениваемой паре, и записываем это число с плюсом. Затем подсчитываем справа же число пар, значение ранга y в которых меньше значения y из оцениваемой пары, и записываем это число рангов со знаком минус. Перед началом оценок каждой пары следует выявить совпадающие ранги (в одноименном ряду) и не принимать в расчет те пары рангов, в которых они встречаются. Первая пара 1—10 имеет ранг $y = 10$, больше которого справа в ряду y имеется 3 ранга: 12, 13, 11. Записываем $+3$. Меньше 10 справа насчитывается 9 рангов. Записываем под оцениваемой парой -9 . Совпадающих рангов нет. Оцененную пару 1—10 исключаем из дальнейших подсчетов. Во второй паре 2—12 ранг $y = 12$, больше которого справа имеется 1 ранг и меньше его 10 рангов, записываем под ним соответственно: $+1$ и -10 . Совпадающих рангов нет.

Оцененную пару и далее каждый раз исключаем из расчетов. В третьей паре 3,5—1 в ряду x имеется ранг 3,5, совпадающий по

значению с рангом $x = 3,5$ в четвертой паре, поэтому четвертую пару исключаем из цикла оценки третьей пары. Получаем: число рангов, больших 1 (справа), равно +9, и меньших 1, равно 0. Также при сравнении четвертой пары; $3,5-3,5$ с другими исключаем одиннадцатую пару: $11-3,5$, так как в ней ранг $y = 3,5$ совпадает с таковым из четвертой пары и т. д.

4. Суммируем положительные и отрицательные суммы оценок:

$$B = +30 + (-45) = -15.$$

5. Находим поправки на объединение рангов в обоих рядах по формуле:

$$B_x \text{ (или } B_y) = \Sigma n(n+1)/2, \quad (2.08)$$

где B_x — поправка на объединение рангов в ряду x ; B_y — поправка на объединение рангов в ряду y ; n — число объединенных рангов в каждой группе.

В ряду x объединено по два ранга в двух группах: 3,5; 3,5; и 8,5; 8,5; отсюда по формуле (2.08):

$$B_x = [2(2-1) + 2(2-1)]/2 = 2.$$

В ряду y объединено два ранга в одной группе: 3,5; 3,5, отсюда:

$$B_y = [2(2-1)]/2 = 1.$$

6. Подставляем в формулу (2.07) найденные значения B , B_x и B_y :

$$\tau = -15 / \sqrt{\left[\frac{13(13-1)}{2} - 2 \right] \left[\frac{13(13-1)}{2} - 1 \right]} = -0,196.$$

7. Оценку значимости коэффициента корреляции Кендэла производим по формуле:

$$t = \tau / \sqrt{\frac{2(2N+5)}{9N(N-1)}} = \frac{\tau}{m_\tau} \geq 3, \quad (2.09)$$

где t — критерий достоверности Стьюдента; τ — коэффициент корреляции Кендэла; N — объем выборки; m_τ — ошибка коэффициента корреляции, которую можно получить из табл. 8П, для $8 < N < 31$. Коэффициент корреляции рангов является значимым, если t равно или превышает 3 или табличное значение (табл. 3П), при числе степеней свободы $\nu = N - 2$.

В нашем примере по формуле (2.09):

$$t = 0,196 / \sqrt{\frac{2(2 \cdot 13 + 5)}{9 \cdot 13(13-1)}} = 0,93$$

ту же величину получаем при $m_\tau = 0,2099$ (табл. 8П).

Вычисленное значение 0,93 значительно меньше табличного 2,20 (табл. 3П) при числе степеней свободы $13-2 = 11$ и доверительном уровне 95%. Следовательно, коэффициент корреляции рангов, равный $-0,196$, является недостоверным, что подтверждает вывод, сделанный в § 2.02, о несовместимости или непригодности специалистов x и y в качестве экспертов.

Формула (2.09) значительно упрощается после возведения ее в квадрат и очевидных преобразований:

$$\frac{N(N-1)\tau^2}{2(2N+5)} \geq 1. \quad (2.10)$$

Если левая часть этого неравенства будет больше единицы, то коэффициент Кендэла значим, а если левая часть неравенства будет меньше единицы, то коэффициент Кендэла будет недостоверным, или при $v(1) = 1$ и $v(2) = N - 2$ значимость коэффициента можно оценить по критерию Фишера (о технике его применения см. § 4.01). Для рассмотренного примера при $N = 13$; $\tau = 0,196$ по формуле (2.10)

$$\frac{13(13-1)0,196^2}{2(2 \cdot 13 + 5)} = 0,09666 < 1.$$

Следовательно, данная величина коэффициента корреляции Кендэла недостоверна. При $P_1 = 95\%$ и $v(1) = 1$ $v(2) = 13 - 2 = 11$, по табл. 9П $F' = 4,84 > 0,09666$, что подтверждает тот же вывод. По формуле

$$\frac{\tau^2(N-1)}{0,4 + 1/N} \geq 10, \quad (2.11)$$

величина коэффициента корреляции Кендэла оценивается наиболее просто. Подставив в нее известные величины, получим:

$$\frac{0,196^2(13-1)}{0,4 + 1/13} = 2,0521,$$

что меньше 10. Таким образом подтверждается уже сделанный вывод о недостоверности данной величины коэффициента корреляции.

Если при вычислении коэффициента Кендэла в рядах x и y объединения рангов не производится (в том случае, когда не имеется совпадающих вариантов в любом из этих рядов), то формула (2.07) обращается в более простую, которая большей частью и применяется:

$$\tau = \frac{2B}{N(N-1)}, \quad (2.12)$$

где обозначения те же, что и к формуле (2.07). Процесс образования суммы рангов B — см. выше в этом же параграфе. Без учета поправок на объединение рангов коэффициент Кендэла по формуле (2.12) равен:

$$\tau = \frac{-2 \cdot 15}{13(13-1)} = -0,192,$$

что несколько отличается от величины, полученной по формуле (2.07). Значения $N(N-1)$ и $2/N(N-1)$ вычислены в табл. 8П для N от 2 до 31. Оценка достоверности коэффициента, полученного по формуле (2.12), производится по той же формуле (2.09).

§ 2.04. Коэффициент корреляции для малых выборок

Коэффициент корреляции вычисляется в тех случаях, когда требуется выразить количественно силу связи между двумя сопряженными признаками, если известно, что зависимость одного из признаков от другого близка к прямолинейной. Если форма связи неизвестна, ее следует предварительно установить по графику (§ 2.01). При взвешенных рядах x (аргумент) и y (функция) применяют метод из § 2.05, а при невзвешенных рядах вычисляется коэффициент корреляции по формуле:

$$r = \frac{\sum xy - (\sum x \cdot \sum y)/N}{\sqrt{[\sum x^2 - (\sum x)^2/N][\sum y^2 - (\sum y)^2/N]}} \quad (2.13)$$

где r — коэффициент корреляции; x — варианты ряда аргумента; y — варианты ряда функции; N — объем выборки (число пар вариантов, или объем одного из рядов x или y).

В конце § 2.05 приведен способ вычисления коэффициента корреляции через отклонения от средней, который также удобен при небольших выборках или машинном счете.

Вычислим по формуле (2.13) силу связи между высотой растений и урожайностью у сои (табл. 2.04). Вычисления производятся в следующем порядке.

1. Возводим в квадрат значения x и y (столбцы 3 и 4).
2. Перемножаем по строкам значения x на y (столбец 5).
3. Суммируем величины столбцов 1—5, получаем: $\sum x = 839$, $\sum y = 262$, $\sum x^2 = 55869$, $\sum y^2 = 5544$, $\sum xy = 16696$.

Таблица 2.04. Вычисление коэффициента корреляции для малых выборок

Высота расте- ний, см, x	Урожай, ц/га, y		y^2	xy
62	19	3844	361	1178
63	20	3969	400	1260
73	12	5329	144	876
82	17	6724	289	1394
66	16	4356	256	1056
44	18	1936	324	792
43	26	1849	676	1118
54	25	2916	625	1350
63	19	3969	361	1197
74	25	5476	625	1850
74	27	5476	729	1998
64	23	4096	529	1472
77	15	5929	225	1155
839	262	55 869	5544	16 696

Таблица 2.04а. К вычислению коэффициента корреляции с преобразованием данных

$a_x =$ $x - A_x$	$a_y =$ $y - A_y$	a_x^2	a_y^2	$a_x a_y$
—2	—4	4	16	+8
—1	—9	1	81	+9
+9	—3	81	9	—27
+18	+21	324	441	+378
+2	—22	4	484	—44
—20	—21	400	441	+420
—21	+14	441	196	—294
—10	+14	100	196	—140
—1	+4	1	16	—4
+10	+13	100	169	+130
+10	+2	100	4	+20
0	—3	0	9	0
+13	—11	169	121	—143
+7	—5	1725	2183	313

4. Возводим в квадрат суммы столбцов 1 и 2:

$$(\Sigma x)^2 = 839^2 = 703921, \quad (\Sigma y)^2 = 262^2 = 68644.$$

5. Перемножаем итог столбца x на итог столбца y :

$$\Sigma x \cdot \Sigma y = 839 \cdot 262 = 219818.$$

6. Подставляем полученные значения в формулу (2.13):

$$r = \frac{16696 - 219818/13}{\sqrt{(55869 - 703921/13)(5544 - 68644/13)}} = -0,316.$$

В зависимости от особенностей исходных данных могут применяться различные способы вычисления коэффициента корреляции для малых выборок. Если известны средние арифметические обоих рядов (M_x, M_y), то формулу (2.13) можно преобразовать:

$$r = \frac{\Sigma yx - M_y \Sigma x}{\sqrt{(\Sigma x^2 - M_x \Sigma x)(\Sigma y^2 - M_y \Sigma y)}}, \quad (2.13a)$$

где обозначения те же; несколько сокращается объем вычислений при пользовании формулой (2.18), где также участвуют величины средних арифметических. При известных коэффициенте регрессии и средних квадратических отклонениях обоих рядов коэффициент корреляции можно получить из формул (3.10) и (3.11), преобразуя их соответственно:

$$r_{xy} = \frac{s_x a_1}{s_y}; \quad r_{xy} = \frac{a_1 s_x c_x}{s_y c_y} \quad (\text{см. § 3.01}).$$

Приблизительные, графические способы определения силы связи редко дают удовлетворительные результаты, поэтому здесь не рассматриваются.

Некоторое сокращение вычислительной работы достигается иногда путем преобразования исходных данных, в которых можно перенести запятую на любое одинаковое в каждом ряду число знаков вправо или влево или вычесть одно и то же число, от этого величина коэффициента корреляции не изменится. Для примера определим силу связи между высотой растений (в см. x) и весом 1000 семян (в г, y) у сои по данным из первых двух столбцов табл. 2.15. Найдем средние арифметические этих рядов: $M_x = 839/13 = 64,5385$; $M_y = 2868/13 = 220,6154$; округлим эти величины: $A_x = 64$; $A_y = 221$, вычтем последние соответственно из всех вариант рядов x и y и произведем следующие действия (табл. 2.04a).

После вычислений первых двух столбцов сделаем проверку:

$$\Sigma a_x = \Sigma x - N \cdot A_x = 839 - 13 \cdot 64 = 7;$$

$$\Sigma a_y = \Sigma y - N A_y = 2868 - 13 \cdot 221 = -5.$$

Найденные суммы подставляем в формулу:

$$r_{xy} = \frac{\Sigma a_x a_y - m_y \Sigma a_x}{\sqrt{(\Sigma a_x^2 - m_x \Sigma a_x)(\Sigma a_y^2 - m_y \Sigma a_y)}}, \quad (2.136)$$

где

$$m_x = \frac{\sum a_x}{N} = \frac{7}{13} = 0,53846;$$

$$m_y = \frac{\sum a_y}{N} = \frac{-5}{13} = -0,38462;$$

$$r_{xy} = \frac{313 + 0,38462 \cdot 7}{\sqrt{(1725 - 0,53846 \cdot 7)(2183 - 0,38462 \cdot 5)}} = \frac{315,69}{\sqrt{3754140}}$$

Под знаком радикала, как это часто бывает, получилось большое число, которое в подобных случаях можно уменьшить, представив его в виде произведения с 10 в четной степени:

$$3754140 = 3,754 \cdot 10^6; \sqrt{10^6} = 10^{6/2} = 10^3,$$

следовательно (после сокращения на $10^3 = 1000$):

$$r_{xy} = \frac{0,31569}{\sqrt{3,754}} = +0,163.$$

Напомним, что извлечение квадратного корня равносильно делению на 2 показателя степени подкоренного выражения.

В некоторых случаях, чтобы не отвлекать внимание на знаки чисел, бывает целесообразно вычесть из всех чисел каждого ряда его минимальную варианту, например в рассматриваемом случае 43 в ряду x и 199 в ряду y . Далее надо вычислить новые средние арифметические для редуцированных данных, получить необходимые суммы, затем можно воспользоваться формулой (2.13а).

Оценка достоверности коэффициента корреляции производится путем преобразования его при помощи табл. 26П, из которой берем $z = 0,33$ для $r = 0,32$.

Ошибка вычисляется по формуле:

$$m_z = 1/\sqrt{N-3}, \quad (2.14)$$

где m_z — ошибка преобразованного коэффициента корреляции; N — объем выборки.

Для рассматриваемого примера

$$m_z = 1/\sqrt{13-3} = 0,316,$$

откуда критерий достоверности Стьюдента:

$$t = z/m_z = 0,33/0,32 = 1,03.$$

Вычисленное значение критерия достоверности меньше табличного (табл. 3П), которое на доверительном уровне 0,95 равно 2,2. Число степеней свободы при оценке преобразованного коэффициента корреляции принимается равным: $\nu = N - 2$. В данном случае коэффициент корреляции оказался недостоверным, что могло произойти по двум причинам: из-за малого объема выборки или ввиду отсутствия исследуемой связи в генеральной совокупности данного объекта. Заметим также, что некоторое снижение величины r может произойти из-за чрезмерно большого округления значений x и y .

В табл. 13П приведены минимальные значения коэффициента корреляции, существенные при данном числе вариант. По этой таблице для суждения о достоверности $r = 0,316 \approx 0,32$ требуется не менее 38 парных вариант, поэтому следует повторить вычисление r на большей выборке, после чего при недостоверном значении коэффициента корреляции можно делать вывод об отсутствии достоверной прямой линейной связи между изучаемыми признаками¹.

§ 2.05. Коэффициент корреляции для больших выборок

Определение силы связи между признаками при взвешенных (см. § 1.01) рядах данных производится путем вычисления коэффициента корреляции с построением корреляционной решетки. Перед вычислением любого показателя связи рекомендуется предварительно построить график данной взаимосвязи для того, чтобы можно было обоснованно выбрать подходящий показатель силы связи в зависимости от степени криволinéйности эмпирической линии регрессии. Если эмпирическая линия регрессии значительно отклоняется от прямой линии, следует применить соответствующий показатель криволinéйнной связи (см. § 2.08—2.12).

Коэффициент корреляции для больших групп вычисляется различными способами, здесь рекомендуется формула

$$r = \frac{\sum f_{xy} a_x a_y - (\sum a_x f_x \cdot \sum a_y f_y) / N}{N \sigma_x \sigma_y}, \quad (2.15)$$

где N — объем выборки; f_{xy} — совместные частоты рядов x и y ; a_x, a_y — отклонения вариант рядов x и y от условных средних: A_x и A_y ; f_x — частоты ряда x ; f_y — частоты ряда y ; σ_x — сигма ряда x ; σ_y — сигма ряда y .

Отклонения вариант в рядах равны:

$$a_x = (x - A_x) / c_x, \quad a_y = (y - A_y) / c_y,$$

где x, y — варианты рядов x и y ; A_x, A_y — центральные варианты (условные средние) рядов x и y ; c_x, c_y — классовые интервалы в рядах x и y .

Сигмы обоих рядов вычисляются по формулам:

$$\sigma_x = \sqrt{\frac{\sum a_x^2 f_x - (\sum a_x f_x)^2 / N}{N}}, \quad \sigma_y = \sqrt{\frac{\sum a_y^2 f_y - (\sum a_y f_y)^2 / N}{N}}, \quad (2.16)$$

где σ_x, σ_y — сигмы рядов x и y ; a_x, a_y — условные отклонения рядов x и y ; f_x, f_y — частоты рядов x и y ; N — объем выборки.

Вычислим коэффициент корреляции для проверки предположения о существовании зависимости веса 1000 зерен от продолжитель-

¹ Дополнение коэффициента корреляции до единицы, т. е. значение выражения: $1 - |r|$, называется коэффициентом алиенации, и позволяет судить о степени независимости изучаемых факторов. В рассмотренном примере коэффициент алиенации равен: $1 - |0,32| = 0,68$.

Таблица 2.05. К вычислению коэффициента корреляции между вегетационным периодом и весом 1000 зерен ячменя

Вес 1000 зерен, г, y	Вегетационный период, дни, x										
	57,5—62,4	62,5—67,4	67,5—72,4	72,5—77,4	77,5—82,4	82,5—87,4	87,5—92,5	a_y	f_y	$a_y f_y$	$a_y^2 f_y$
	60	65	70	75 = A_x	80	85	90				
55—61 58		1 —6	2 —6	3 0	7 21	1 6		+3	14	42	126
49—54,9 52		5 —20	8 —16	10 0	11 22	2 8	2 12	+2	38	76	152
43—48,9 46		7 —14	10 —10	17 0	33 33	3 6	1 3	+1	71	71	71
37—42,9 40 = A_y	1 0	8 0	9 0	20 0	18 0	1 0	1 0	0	58	0	0
31—36,9 34	1 3	4 8	5 5	10 0	5 —5	1 —2		—1	26	—26	26
25—30,9 28		3 12		3 0				—2	6	—12	24
19—24,9 22		1 6						—3	1	—3	9
a_x	—3	—2	—1	0	+1	+2	+3	$\Sigma = 148$			
f_x	2	29	34	63	74	8	4	$N = 214$			
$a_x f_x$	—6	—58	—34	0	74	16	12	$\Sigma = 4$			
$a_x^2 f_x$	18	116	34	0	74	32	36	$\Sigma = 310$			
$f_x a_x a_y$	3	—14	—27	0	71	18	15	$\Sigma = 66$			

ности вегетационного периода у сортов ячменя (табл. 2.05). Действия производятся в следующем порядке.

1. Массивы данных по обоим признакам разбиваются на классы (см. § 1.01) и производится разноска вариантов совместно обоих рядов по клеткам корреляционной таблицы (см. § 2.01). Число классов в рядах x и y может быть различным. Аргумент (x) располагается в горизонтальном ряду таблицы слева направо. Значения функции (y) возрастают снизу вверх (они могут и, напротив, возрастать сверху вниз; от этого результат не изменится, но будет труднее представить форму зависимости по расположению частот в таблице).

2. Вычисляется эмпирическая линия регрессии (см. § 2.01) для выяснения степени криволинейности данной связи. Если эмпирическая линия регрессии близка к прямой линии и не является двускатной (см. рис. 15), можно вычислять коэффициент корреляции, в противном случае вычисляется показатель корреляции для криволинейных связей (§ 2.06—2.12).

3. В корреляционной таблице произвольно выбирается условный центр. Обычно это близкая к центру клетка, содержащая наибольшую частоту (в данном случае — клетка с частотой 20). В ряду x против нее будет находиться условная средняя $A_x = 75$, а в ряду y средняя $A_y = 40$.

4. Вычислим условные отклонения a_x :

$$(60-75)/5 = -3; \quad (65-75)/5 = -2; \quad (70-75)/5 = -1;$$

$$(75-75)/5 = 0; \quad (80-75)/5 = +1 \text{ и т. д.,}$$

записывая значения a_x в горизонтальной строке.

5. Таким же образом, от условной средней $A_y = 40$, вычисляются отклонения a_y , которые записываются в вертикальной графе табл. 2.05.

6. Условные отклонения перемножаем на соответствующие им частоты: $a_x f_x = -3 \cdot 2 = -6$; $-2 \cdot 29 = -58$; $-1 \cdot 34 = -34$ и т. д.; $a_y f_y = +3 \cdot 14 = 42$; $+2 \cdot 38 = 76$ и т. д.

7. Квадраты условных отклонений перемножаем на соответствующие им частоты: $a_x^2 f_x = (-3)^2 \cdot 2 = 18$; $(-2)^2 \cdot 29 = 116$; $(-1)^2 \cdot 34 = 34$ и т. д.

Также для ряда y : $a_y^2 f_y = 3^2 \cdot 14 = 126$; $2^2 \cdot 38 = 152$; $1^2 \cdot 71 = 71$; $0^2 \cdot 58 = 0$; $(-1)^2 \cdot 26 = 26$; $(-2)^2 \cdot 6 = 24$; $(-3)^2 \cdot 1 = 9$.

8. Находим произведения условных отклонений между собой и на имеющуюся совместную для них частоту. Начнем перемножать эти числа по столбцам слева направо;

$$\text{первый столбец: } a_x a_y f_{xy} = (-3) \cdot 0 \cdot 1 = 0; \quad (-3) \cdot 1 \cdot (-1) = 3;$$

второй столбец: $(-2) \cdot 3 \cdot 1 = -6$; $(-2) \cdot 2 \cdot 5 = -20$; $(-2) \cdot 1 \cdot 7 = -14$ и т. д., записывая полученные произведения в правом верхнем углу соответствующих клеток таблицы.

9. Суммируем совместные частоты по столбцам и записываем в горизонтальной строке снизу таблицы: 3, -14, -27, 0, 71, 18, 15.

10. Получаем суммы по строкам (ряд x) и по столбцам (ряд y):
 $\Sigma f_x = N = 214$; $\Sigma a_x f_x = 4$; $\Sigma f_y = N = 214$;

$$\Sigma a_x^2 f_x = 310; \quad \Sigma f_{xy} a_x a_y = 66;$$

$$\Sigma a_y f_y = 148; \quad \Sigma a_y^2 f_y = 408.$$

Для проверки вычисление этих сумм следует повторить.

11. Находим сигму ряда x и сигму ряда y по формулам (2.16):

$$\sigma_x = \sqrt{\frac{310 - 4^2/214}{214}} = 1,203, \quad \sigma_y = \sqrt{\frac{408 - 148^2/214}{214}} = 1,195.$$

12. Подставляя найденные выше суммы и величины сигм в формулу (2.15), получаем:

$$r = \frac{66 - 4 \cdot 148/214}{214 \cdot 1,203 \cdot 1,195} = 0,2055.$$

Ошибку коэффициента корреляции вычисляют по двум формулам в зависимости от численности совокупности. Для небольших групп данных при $N < 100$ применяют преобразование через z и ошибку коэффициента корреляции находят по формуле (2.14), а для больших групп при $N > 100$ ошибку можно определить по формуле

$$m_r = (1 - r^2)/\sqrt{N}, \quad (2.17)$$

где m_r — ошибка коэффициента корреляции; r — коэффициент корреляции; N — объем выборки ($N > 100$).

Для рассмотренного примера по формуле (2.17)

$$m_r = (1 - 0,2055^2)/\sqrt{214} = 0,065.$$

Достоверность найденного коэффициента корреляции определяется по критерию Стьюдента:

$$t = r/m_r = 0,2055/0,065 = 3,16.$$

Число степеней свободы при определении значимости коэффициента корреляции равно: $\nu = N - 2 = 214 - 2 = 212$. По табл. 3П при $P'_1 = 0,95$ $t = 2,0$.

Вычисленное выше значение, равное 3,16, больше табличного. Следовательно, между продолжительностью вегетационного периода и весом 1000 семян, судя по выборке из 214 сортов ячменя, имеется положительная, небольшая, но достоверная зависимость: чем продолжительнее вегетационный период, тем больше вес 1000 зерен. Оценку коэффициента корреляции можно произвести также по табл. 13П (см. § 2.04).

При расчете коэффициента корреляции на ЭВМ как для малых, так и для больших рядов (без группирования в классы) удобна формула:

$$r = \frac{\Sigma (x - M_x)(y - M_y)}{\sqrt{\Sigma (x - M_x)^2 \Sigma (y - M_y)^2}}, \quad (2.18)$$

где r — коэффициент корреляции; $x - M_x$ — отклонения значений ряда аргумента от его средней (M_x); $y - M_y$ — отклонения значений ряда функции от ее средней (M_y).

Пример расчета по этой формуле приведен в § 3.01.

§ 2.06. Совместное вычисление коэффициента корреляции и прямого корреляционного отношения между взвешенными рядами

Данный метод может быть применен главным образом тогда, когда при наличии большой выборки данных требуется количественно определить не только силу связи между изучаемыми признаками, но и степень ее криволинейности. Такие случаи могут встретиться, если по эмпирической линии регрессии нельзя определенно судить, насколько изучаемая связь отклоняется от прямолинейной (например, см. рис. 14) или, если требуется помимо количественного выражения силы связи, получить также и представление о тенденции (знаке) данной зависимости. Квадрат коэффициента корреляции, который вычисляется и оценивается здесь, иногда называют коэффициентом детерминации. Рассматриваемый метод включает также и вычисление точек эмпирической линии регрессии y/x , которыми можно воспользоваться для построения графика изучаемой взаимосвязи.

Вычисление квадратов коэффициента корреляции и прямого корреляционного отношения по совмещенному алгоритму ведется по формулам:

$$r = \frac{\sum (a_x \sum f a_y) - (\sum n_x a_x \sum n_y a_y) / N}{\sqrt{[\sum n_x a_x^2 - (\sum n_x a_x)^2 / N] [\sum n_y a_y^2 - (\sum n_y a_y)^2 / N]}}, \quad (2.19)$$

$$\eta_{y/x}^2 = \frac{\sum \frac{(\sum f a_y)^2}{n_x} - (\sum n_y a_y)^2 / N}{\sum n_y a_y^2 - (\sum n_y a_y)^2 / N}, \quad (2.20)$$

где r — коэффициент корреляции; a_x — условные отклонения вариант ряда x от начала ряда; f — частоты корреляционной решетки; a_y — условные отклонения вариант ряда y от начала ряда; n_x — суммы частот по столбцам корреляционной решетки; n_y — суммы частот по строкам корреляционной решетки; N — объем выборки; $\eta_{y/x}$ — прямое корреляционное отношение.

Последовательность вычислений по формулам (2.19) и (2.20) следующая.

1. Составляем корреляционную решетку таким образом, чтобы варианты ряда аргумента (x) располагались в верхней строке решетки слева направо, а варианты ряда функции (y) располагались бы в вертикальном столбце снизу вверх (см. табл. 2.06).

2. Суммируем частоты по столбцам (n_x) и по строкам (n_y).

3. Записываем условные отклонения a_x и a_y от начала рядов x и y в виде натурального ряда чисел, начиная с 0, соответственно направлению возрастания вариант в обоих рядах.

Таблица 2.06. Полный анализ корреляционной зависимости веса семян (y) от продолжительности вегетационного периода (x) у ячменя

y	x	60	65	70	75	80	85	90	n_y	a_y
58			1	2	3	7	1		14	6
52			5	8	10	11	2	2	38	5
46			7	10	17	33	3	1	71	4
40	1		8	9	20	18	1	1	58	3
34	1		4	5	10	5	1		26	2
28			3		3				6	1
22			1						1	0
$\sum^{n_x} f a_y$	2	29		34	63	74	8	4	$N = 214$	790
$(\sum f a_y)^2 / n_x$	5	94		129	219	293	33	17	790	3222
$\sum^{a_x} f a_y$	12,50	304,69		489,44	761,29	1160,12	136,13	72,25	2936,42	
$(\sum f a_y) / n_x$	0	1		2	3	4	5	6	646	2260
y/x	2,50	3,24		3,79	3,48	3,96	4,13	4,25		
	37,00	41,44		44,74	42,88	45,76	46,78	47,50		

4. Получаем по столбцам суммы произведений частот на соответствующие им условные отклонения a_y : первая сумма $5 = 1 \cdot 3 + 1 \cdot 2$; последняя сумма $17 = 2 \cdot 5 + 1 \cdot 4 + 1 \cdot 3$. В итоге этой строки $\Sigma fa_y = 790$.

5. Значения Σfa_y возводим в квадрат и делим каждое из них на соответствующие им суммы частот $(\Sigma fa_y)^2/n_x$, получаем: 12,50; 304,69; 489,44 и т. д., в итоге этой строки:

$$\sum \frac{(\Sigma fa_y)^2}{n_x} = 2936,42.$$

6. Для получения точек эмпирической линии регрессии отдельные значения Σfa_y : 5,94; 129 и т. д. надо разделить на соответствующие им n_x : 2, 29, 34 и т. д., получим $(\Sigma fa_y)/n_x$: 2,50; 3,24; 3,79 и т. д. Затем каждое из полученных значений умножается на величину классового интервала $c = 6$ и к каждому произведению прибавляется минимальная варианта ряда y (она в табл. 2.06 равна 22): $2,50 \cdot 6 + 22 = 37,00$; $3,24 \cdot 6 + 22 = 41,44$; $3,79 \cdot 6 + 22 = 44,74$ и т. д. Полученные значения y/x представляют собой точки эмпирической линии регрессии.

7. Перемножаем все значения n на соответствующие условные отклонения a_x и складываем произведения: $2 \cdot 0 + 29 \cdot 1 + 34 \cdot 2 + 63 \cdot 3 + 4 \cdot 6 = 646 = \Sigma n_x a_x$.

8. Получаем сумму $\Sigma n_x a_x^2 = 2260$, перемножая значения n_x на квадраты условных отклонений: $2 \cdot 0^2 + 29 \cdot 1^2 + 34 \cdot 2^2 + 63 \cdot 3^2$ и т. д.

9. Подобно действиям в пунктах 7 и 8, получаем суммы:

$$\Sigma n_y a_y = 790 = 1 \cdot 0 + 6 \cdot 1 + 26 \cdot 2 \quad \text{и т. д.};$$

$$\Sigma n_y a_y^2 = 3222 = 1 \cdot 0^2 + 6 \cdot 1^2 + 26 \cdot 2^2 + 58 \cdot 3^2 \text{ и т. д.}$$

10. Получаем сумму $\Sigma (a_x \Sigma fa_y) = 2448 = 0 \cdot 5 + 1 \cdot 94 + 2 \cdot 129 + 3 \cdot 219 + 4 \cdot 293 + 5 \cdot 33 + 6 \cdot 17$.

11. Подставляем найденные суммы в формулы (2.19) и (2.20):

$$r = \frac{2448 - \frac{646 \cdot 790}{214}}{\sqrt{\left(2260 - \frac{646^2}{214}\right) \left(3222 - \frac{790^2}{214}\right)}} = +0,205; \quad r^2 = 0,0422;$$

$$\eta_{y/x}^2 = \frac{2936,42 - \frac{790^2}{214}}{3222 - \frac{790^2}{214}} = 0,0656; \quad \eta_{y/x} = 0,256.$$

12. Оценку найденных показателей производим по критерию Фишера.

Достоверность коэффициента детерминации (квадрата коэффициента корреляции) определяется по формуле

$$F = \frac{r^2(N-2)}{1-r^2}, \quad (2.21)$$

где F — критерий Фишера; r — коэффициент корреляции; N — объем выборки.

По формуле (2.21):

$$F = \frac{0,0422 (214 - 2)}{1 - 0,0422} = 9,34.$$

Числа степеней свободы при оценке коэффициента детерминации принимаются: $\nu(1) = 1$, а $\nu(2) = N - 2 = 214 - 2 = 212$; отсюда значение критерия Фишера по табл. 9П при $P_1 = 95\%$ равно 3,89. Вычисленное по формуле (2.21) значение критерия больше табличного, поэтому делаем вывод о существовании достоверной прямой зависимости веса 1000 зерен от продолжительности вегетационного периода у ячменя.

Достоверность квадрата корреляционного отношения определяется по формуле

$$F = \frac{\eta_{y/x}^2 (N - k_x)}{(1 - \eta_{y/x}^2) (k_x - 1)}, \quad (2.22)$$

где F — критерий Фишера; $\eta_{y/x}$ — прямое корреляционное отношение; N — объем выборки; k_x — число классов в ряду x .

По формуле (2.22):

$$F = \frac{0,0656 (214 - 7)}{(1 - 0,0656) (7 - 1)} = 2,42.$$

Числа степеней свободы при определении достоверности $\eta_{y/x}^2$ по табл. 9П принимаются равными числу классов ряда x без единицы:

$$\nu(1) = k_x - 1 = 7 - 1 = 6,$$

и объему выборки минус число классов ряда x :

$$\nu(2) = N - k_x = 214 - 7 = 207.$$

Отсюда при $P_1 = 95\%$ находим F' табличное, равное 2,14. Вычисленное по формуле (2.22) значение критерия составляет 2,42, т. е. больше табличного, поэтому делаем вывод о достоверности также и криволинейной зависимости веса 1000 зерен от продолжительности вегетационного периода у ячменя.

§ 2.07. Критерии криволинейности

В биологических исследованиях нередко возникает вопрос, к какому типу отнести изучаемую связь: к прямой или криволинейной. Оценка степени криволинейности корреляционной зависимости производится при помощи критериев криволинейности, из которых здесь рассмотрено пять в порядке возрастания трудоемкости их вычисления:

$$a) \eta_{y/x}^2 - r^2 \geq 0,1, \quad (2.23)$$

где $\eta_{y/x}$ — прямое корреляционное отношение; r — коэффициент корреляции.

Связь считается криволинейной, если разность квадратов $\eta_{y/x}$ и r больше 0,1. Для рассмотренного в предыдущем параграфе примера:

$$0,0656 - 0,0422 = 0,02 < 0,1,$$

поэтому связь между продолжительностью вегетационного периода и весом 1000 зерен практически можно считать прямолинейной. Кроме того, прямолинейная связь сравнительно слаба и вычисление корреляционного отношения лишь немного повысило ее оценку.

$$\text{б) } N (\eta_{y/x}^2 - r^2) \geq 11,37, \quad (2.24)$$

где N — объем выборки; $\eta_{y/x}$ — прямое корреляционное отношение; r — коэффициент корреляции; 11,37 — критерий криволинейности Блэкмана.

Исходя из данных предыдущего примера (а), по формуле Блэкмана

$$214 (0,0656 - 0,0422) = 5,03.$$

Полученная разница (5,03) меньше 11,37, поэтому считаем, что рассматриваемая связь может считаться близкой к прямолинейной.

$$\text{в) } F = \frac{(\eta_{y/x}^2 - r^2) (N - k_x)}{(1 - \eta_{y/x}^2) (k_x - 2)}, \quad (2.25)$$

где F — критерий Фишера; $\eta_{y/x}$ — прямое корреляционное отношение; r — коэффициент корреляции; N — объем выборки; k_x — число классов в ряду x .

Для рассматриваемого примера о ячмене по формуле (2.25):

$$F = \frac{(0,0656 - 0,0422) (214 - 7)}{(1 - 0,0656) (7 - 2)} = 1,04.$$

Числа степеней свободы при определении значимости F равны числу классов ряда x без двух:

$$\nu (1) = k_x - 2 = 7 - 2 = 5,$$

и объему выборки минус число классов ряда x :

$$\nu (2) = N - k_x = 214 - 7 = 207.$$

По табл. 9П при $P_1 = 95\%$ на пересечении строк $\nu (1) = 5$ и $\nu (2) = 207$ находим $F' = 2,26$. Величина критерия, вычисленная по формуле (2.22), меньше табличной: $1,04 < 2,26$, поэтому делаем вывод о несущественном отличии рассматриваемой связи от прямолинейной.

г) оценка степени криволинейности связи при помощи критерия Стьюдента производится по формуле

$$t = 0,5 \sqrt{\frac{N}{(\eta_{y/x}^2 - r^2)^{-1} - 2 + \eta_{y/x}^2 + r^2}} \geq 3, \quad (2.26)$$

или (что одно и тоже) по формуле:

$$t^2 = \frac{0,25N}{(\eta_{y/x}^2 - r^2)^{-1} - 2 + \eta_{y/x}^2 + r^2} \geq 9, \quad (2.26a)$$

где t — критерий Стьюдента; $\eta_{y/x}$ — прямое корреляционное отношение; r — коэффициент корреляции; N — объем выборки.

Для примера, рассматриваемого в данном параграфе, по формуле (2.26)

$$t = 0,5 \sqrt{\frac{214}{(0,0656 - 0,0422)^{-1} - 2 + 0,0656 + 0,0422}} = 1,15.$$

Вычисленное значение меньше 3 и меньше табличного (табл. 3П) при $P_1 = 95\%$, которое равно 1,960; поэтому полагаем, что рассматриваемая связь несущественно отклоняется от прямолинейной.

д) оценка степени криволинейности по критерию Романовского:

$$\frac{\frac{N - k - 2}{k - 2} \cdot \frac{\eta_{y/x}^2 - r^2}{1 - \eta_{y/x}^2} - 1}{\sqrt{\frac{2(N - 4)}{(k - 2)(N - k - 4)}}} \geq 3, \quad (2.27)$$

где N — объем выборки; k — число групп, на которое разбивался ряд при вычислении корреляционного отношения; $\eta_{y/x}$ — прямое корреляционное отношение; r — коэффициент корреляции.

Для рассматриваемого примера по критерию Романовского также подтверждается уже сделанный вывод:

$$\frac{\frac{214 - 7 - 2}{7 - 2} \cdot \frac{0,256^2 - 0,205^2}{1 - 0,256^2} - 1}{\sqrt{\frac{2(214 - 4)}{(7 - 2)(214 - 7 - 4)}}} = 0,05 < 3.$$

Из пяти рассмотренных критериев криволинейности первые два просты по вычислению, но менее точны, чем три последних. Заметим, что для решения вопроса о степени криволинейности связи во многих случаях можно удовлетвориться визуальным анализом графика связи по эмпирическим данным.

§ 2.08. Прямое корреляционное отношение для невзвешенных рядов

Корреляционные отношения дисперсий, прямое и обратное, служат для оценки силы связи между двумя признаками, когда последние могут находиться не только в прямолинейной, но и в криволинейной зависимости между собой. При более детальном исследовании вместо корреляционных отношений лучше применять методы дисперсионного анализа (см. главу 4), с которыми они тесно связаны.

Схема вычисления корреляционных отношений (§ 2.06, 2.08—2.11) приводится в основном по алгоритму Н. А. Плохинского (1970), как наиболее распространенному к настоящему времени. Су-

ществуют другие, более строго теоретически обоснованные схемы расчета корреляционных отношений [см., например (Урбах, 1964)], которые, однако, более трудоемки по технике вычислений, так как требуют выполнения однофакторного дисперсионного анализа, при помощи которого также по этим схемам производится и оценка значимости полученных показателей связи. Между тем сравнительно просто вычисляемый показатель криволинейной связи для невзвешенных рядов, имеющий способ для непосредственной оценки его значимости и сравнимый с коэффициентом корреляции, в биологических исследованиях необходим, что и делает пока допустимым использование приводимых ниже схем расчета корреляционных отношений.

Под прямым корреляционным отношением понимается такой показатель, который измеряет силу криволинейной или прямолинейной зависимости признака y (функция) от признака x (аргумент), его расчет производится по формуле

$$\eta_{y/x} = \sqrt{\frac{n \sum y/x - M_y)^2}{\sum (y - M_y)^2}}, \quad (2.28)$$

где $\eta_{y/x}$ — прямое корреляционное отношение; y/x — значения (точки) эмпирической линии регрессии; M_y — средняя арифметическая ряда y ; y — варианты ряда; n — число вариантов в частной группе.

Вычисление прямого корреляционного отношения покажем на примере анализа взаимосвязи между числом вегетативных и фертильных побегов у костра безостого (табл. 2.07). Цель анализа: выяснить направление взаимосвязи между побегами, т. е. какая группа побегов больше влияет на численность другой. Предположим, что число вегетативных побегов обуславливает число фертильных, поэтому примем число первых за аргумент (x), а число вторых — за функцию (y) и соответственно вычислим прямое корреляционное отношение.

1. Точки эмпирической линии регрессии y/x по невзвешенному ряду получаем путем усреднения по группам. В данном случае эти группы образовались естественно: по 3 растения каждого сорта ($n = 3$); в других случаях группы составляют искусственно, не менее 2 вариантов в каждой. Суммы: $164 + 184 + 185 = 533$, $180 + 183 + 205 = 568$ и т. д. (столбец 4) делим на число вариантов в группе (здесь оно везде равно 3) и получаем точки эмпирической линии регрессии y/x : 177,7; 189,3; 110,0 и т. д. (столбец 5).

2. Вычисляем среднюю арифметическую ряда y :

$$M_y = \frac{\sum y}{N} = \frac{4397}{27} = 162,85.$$

3. Вычисляем квадраты отклонений точек эмпирической линии регрессии от средней арифметической (столбцы 6 и 7):

$$\sum (y/x - M_y)^2 = 7783,3225.$$

4. Сумму квадратов отклонений вариантов получаем возведением в квадрат разницы каждой варианты y со средней арифметической

Таблица 2.07. К вычислению прямого корреляционного отношения

Форма костра безостого	Число побегов		Суммы y , (по 3 растения)	y/x	$y/x - M_y$	$(y/x - M_y)^2$
	вегетативных, x	фертильных, y				
1	132	164	533	177,7	14,84	220,5225
	164	184				
	184	185				
2	76	180	568	189,3	26,45	699,6025
	136	183				
	96	205				
3	257	123	330	110,0	—52,85	2793,1225
	78	92				
	143	115				
4	270	114	564	188,0	25,15	632,5225
	146	250				
	121	200				
5	206	150	475	158,3	—4,55	20,7025
	197	142				
	227	183				
6	53	80	364	121,3	—41,55	1726,4025
	92	124				
	161	160				
7	147	175	604	201,3	38,45	1478,4025
	102	209				
	106	220				
8	170	152	509	169,7	6,85	46,9225
	107	179				
	155	178				
9	20	126	450	150,0	—12,85	165,1225
	27	146				
	36	178				
$N = 27$	3609	4397	4397		—0,05	7783,3225

и одновременным суммированием квадратов на счетах или счетной машине. Квадраты чисел можно также получить по таблицам.

$$\Sigma (y - M_y)^2 = 41087.$$

5. По формуле (2.28)

$$\eta_{y/x} = \sqrt{\frac{3 \cdot 7783,3225}{41087}} = 0,75.$$

6. Ошибку квадрата корреляционного отношения, в том числе и обратного, вычислим по формуле:

$$m_{\eta^2} = (1 - \eta^2) \frac{v_1}{v_2}, \quad (2.29)$$

где m_{η^2} — ошибка корреляционного отношения; η — корреляционное отношение; ν_1, ν_2 — числа степеней свободы: $\nu_1 = k - 1$, $\nu_2 = N - k$; k — число групп (классов), на которое разделен ряд; N — объем выборки.

$$\nu_1 = 9 - 1 = 8; \quad \nu_2 = 27 - 9 = 18.$$

$$m_{\eta^2} = (1 - 0,75^2) (8/18) = 0,19.$$

Оценку достоверности квадрата корреляционного отношения делают по формуле

$$F = \frac{\eta^2}{m_{\eta^2}} \geq F', \quad (2.30)$$

где F — вычисленное значение критерия Фишера; F' — его табличное значение; η — корреляционное отношение; m_{η^2} — ошибка квадрата корреляционного отношения.

$$F = \frac{0,75^2}{0,19} = 2,89.$$

По табл. 9П при $P'_1 = 0,95$ и $\nu(1) = \nu_1 = 8$; $\nu(2) = \nu_2 = 18$ $F' = 2,51$; следовательно, на данном доверительном уровне число вегетативных побегов у костра безостого достоверно влияет на число фертильных побегов.

7. Оценку достоверности прямого и обратного корреляционных отношений удобно, особенно при отсутствии табл. 9П и при счете на ЭВМ, делать по критериям:

$$\sqrt{\frac{\frac{(\nu_2 - 2) \eta^2}{(1 - \eta^2) \nu_1} - 1}{\frac{2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)}}} \geq 3, \quad (2.31)$$

или:

$$\frac{\left(\frac{\nu_2 - 2}{1 - \eta^2} \cdot \frac{\eta^2}{\nu_1} - 1 \right)^2 \nu_1 (\nu_2 - 4)}{2(\nu_1 + \nu_2 - 2)} \geq 9, \quad (2.32)$$

где $\nu_1 = k - 1$; $\nu_2 = N - k$; k — число групп, на которое был разделен ряд при вычислении корреляционного отношения; N — объем выборки; η — корреляционное отношение.

Как видим из формул, число степеней свободы ν_2 должно быть не меньше 5. При величине критерия больше или равной 3 (или 9, по второму варианту формулы) корреляционное отношение будет достоверным. Уровень достоверности при таком значении критерия примерно $P'_2 = 0,99$. Для менее ответственных выводов можно принимать за его критическое значение соответственно 2,5, или 7,5.

Вычислим критерий достоверности (2.31) для прямого корреляционного отношения: $\nu_1 = 9 - 1 = 8$; $\nu_2 = 27 - 9 = 18$:

$$\sqrt{\frac{[(18 - 2) 0,75^2] / [(1 - 0,75^2) 8] - 1}{\frac{2(8 + 18 - 2)}{8(18 - 4)}}} = 2,40.$$

Критерий здесь меньше 3, следовательно, на уровне достоверности больше 0,95 данное корреляционное отношение недостоверно. Для более полного представления о направлении этой связи, а также о том, какой признак действительно является аргументом в ней, требуется вычислить обратное корреляционное отношение, что и сделано в следующем параграфе.

§ 2.09. Обратное корреляционное отношение для невзвешенных рядов

В тех случаях, когда после вычисления прямого корреляционного отношения остается неуверенность в том, что аргумент в парной взаимосвязи был выбран верно, необходимо вычислить обратное корреляционное отношение, при помощи которого можно оценить обратную связь между признаками, т. е. силу влияния y на x . Схема расчета остается точно такой же, однако все вычисления проводятся с вариантами из ряда x :

$$\eta_{x/y} = \sqrt{\frac{n \sum (x/y - M_x)^2}{\sum (x - M_x)^2}}, \quad (2.33)$$

где $\eta_{x/y}$ — обратное корреляционное отношение; x/\bar{y} — значения (точки) эмпирической линии регрессии; M_x — средняя арифметическая вариант; n — число вариант в частной группе.

1. Получаем точки эмпирической линии регрессии x/y . Для этого суммируем по каждой форме костра варианты x , которые приведены в табл. 2.07 столбец 2 и записываем их в первый столбец табл. 2.08: 480, 308, 478 и т. д.

Таблица 2.08. К вычислению обратного корреляционного отношения для невзвешенных рядов

Групповые суммы ряда, x	x/y	$x/y - M_x$	$(x/y - M_x)^2$	Групповые суммы ряда, x	x/y	$x/y - M_x$	$(x/y - M_x)^2$
480	160,0	26,33	693,2689	306	102,0	—31,67	1 002,9889
308	102,7	—30,97	959,1409	355	118,3	—15,37	236,2369
478	159,3	25,63	656,8969	432	144,0	10,33	106,7089
537	179,0	45,33	2054,8089	83	27,7	—105,97	11 229,6409
630	210,0	76,33	5826,2689				
				3609		—0,03	22 765,9601

2. Каждую групповую сумму делим на число ее составляющих, т. е. на $n = 3$, и получаем групповые средние, или точки эмпирической линии регрессии x/y : 160,0; 102,7; 159,3 и т. д.

3. Средняя арифметическая ряда аргумента равна:

$$M_x = \frac{\sum x}{N} = \frac{3609}{27} = 133,67.$$

4. Сумма квадратов отклонений значений эмпирической линии регрессии x/y от средней арифметической равна (столбцы 3, 4 табл. 2.08):

$$\Sigma (x/y - M_x)^2 = 22765,9601.$$

5. Сумма квадратов отклонений вариант от средней арифметической (получена на счетной машине):

$$\Sigma (x - M_x)^2 = 110720.$$

6. Обратное корреляционное отношение по формуле (2.33):

$$\eta_{x/y} = \sqrt{\frac{3 \cdot 22765,9601}{110720}} = 0,79.$$

7. Оценка достоверности по критерию (2.31):

$$\frac{\frac{(18-2)}{(1-0,79^2)} \cdot \frac{0,79^2}{8} - 1}{\sqrt{\frac{2(8+18-2)}{8(18-4)}}} = 3,55 > 3;$$

и по критерию Фишера (2.29), (2.30):

$$v_1 = 9-1 = 8; \quad v_2 = 27-9 = 18;$$

$$m_{\eta^2} = (1 - 0,79^2) \cdot \frac{8}{18} = 0,17;$$

$$F = \frac{0,79^2}{0,17} = 3,67; \quad F' = 2,51 \quad (\text{при } P'_1 = 0,95) \quad (\text{табл. 9П})$$

Следовательно, обратное корреляционное отношение более достоверно и по величине оказалось больше, чем прямое корреляционное отношение. Поэтому делаем вывод о том, что за ряд аргумента следует принять не число вегетативных побегов, а число фертильных, т. е. число фертильных побегов более влияет на число вегетативных, чем наоборот.

При пользовании методами из § 2.08 и 2.09 следует иметь в виду для сравнимости результатов, что на величину корреляционного отношения влияет способ разбивки рядов на группы (число групп и число вариант, составляющих группы).

§ 2.10. Прямое корреляционное отношение для взвешенных рядов

Прямое корреляционное отношение по данному методу вычисляется в тех случаях, когда исходные данные сгруппированы во взвешенные ряды. Перед началом работы по вычислению отношения требуется установить, что является аргументом (x) и функцией (y) в изучаемой зависимости. Прямое корреляционное отношение для взвешенных

рядов вычисляется по формуле:

$$\eta_{y/x} = \sqrt{\frac{\sum (y/x - M_y)^2 f_x}{\sum (y - M_y)^2 f_y}}, \quad (2.34)$$

где $\eta_{y/x}$ — прямое корреляционное отношение; M_y — средняя арифметическая ряда y .

Остальные обозначения по их смыслу и расположению соответствуют таковым в табл. 2.01, по данным которой оценим зависимость веса 1000 семян от периода созревания у ячменя. Вычисления производятся в следующем порядке.

1. Составляем корреляционную решетку и рассчитываем эмпирическую линию регрессии y/x , точки которой уже вычислены в табл. 2.01.

Таблица 2.09. К вычислению прямого корреляционного отношения для взвешенных рядов

y	f_y	$y f_y$	y/x	f_x	$(y - M_y)^2 f_x$	$(y/x - M_y)^2 f_x$
1	2	3	4	5	6	7
58	14	812	37,0	2	2685,5	102,25
52	38	1976	41,4	29	2341,7	219,31
46	71	3266	44,8	34	243,00	14,365
40	58	2320	42,9	63	998,91	98,438
34	26	884	45,8	74	2678,6	201,47
28	6	168	46,8	8	1564,9	56,180
22	1	22	47,5	4	490,62	44,890
	214	9448	—	214	11 003	736,90

2. Определяем среднюю арифметическую ряда y . Выполнив действия, приведенные в столбце 3 табл. 2.09, получаем:

$$M_y = 9448/214 = 44,15 \text{ г.}$$

3. Сумма квадратов отклонений, умноженных на частоты для ряда y , равна: $\sum (y - M_y)^2 f_y = 11003$ (столбец 6).

4. Сумма квадратов отклонений, умноженных на частоты для эмпирической линии регрессии, равна:

$$\sum (y/x - M_y)^2 f_x = 736,90.$$

Обе последние суммы и слагаемые для них получены на ЭВМ, округление на которой ведется непрерывно в процессе счета до пятизначных цифр.

5. По формуле (2.34):

$$\eta_{y/x} = \sqrt{\frac{736,90}{11003}} = 0,26.$$

6. Ошибка квадрата корреляционного отношения (2.29):

$$m_{\eta^2} = (1 - 0,26^2) (6/207) = 0,027.$$

$$v_1 = 7-1; \quad v_2 = 214-7.$$

Оценка достоверности по критерию Фишера (2.30):

$$F = 0,26^2/0,027 = 2,50; \quad F' = 2,14 \text{ (табл. 9П).}$$

7. Критерий достоверности (2.31) равен:

$$\frac{[(207 - 2) 0,26^2]/[(1 - 0,26^2) 6] - 1}{\sqrt{[2(6 + 207 - 2)]/[6(207 - 4)]}} = 2,51.$$

Оба критерия позволяют заключить, что продолжительность вегетационного периода у ячменя лишь на 95% доверительном уровне достоверно влияет на вес 1000 семян; при более строгой оценке эта связь является недостоверной.

§ 2.11. Обратное корреляционное отношение для взвешенных рядов

Обратное корреляционное отношение для больших классифицированных выборок вычисляется в тех же случаях, в которых оно применяется для невзвешенных рядов (см. § 2.09). Используя пример, рассмотренный в § 2.10, рассчитаем обратное корреляционное отношение, оценивающее зависимость продолжительности вегетационного периода от веса 1000 семян у ячменя (табл. 2.01) по формуле

$$\eta_{x/y} = \sqrt{\frac{\sum (x/y - M_x)^2 f_y}{\sum (x - M_x)^2 f_x}}, \quad (2.35)$$

где $\eta_{x/y}$ — обратное корреляционное отношение; x/y — точки эмпирической линии регрессии; M_x — средняя арифметическая ряда x ; f_x, f_y — частоты рядов.

Порядок вычислений тот же, что и в § 2.10, но с другим рядом (x).

1. Получаем значения эмпирической линии регрессии x/y (столбец 4 табл. 2.10).

2. Средняя арифметическая ряда x :

$$M_x = 16070/214 = 75,094 \text{ (столбцы 1, 2, 3 табл. 2.10).}$$

3. Суммы квадратов отклонений вариант ряда x (столбец 6) и значений эмпирической линии регрессии (столбец 7), помноженных на соответствующие частоты, равны: $\sum (x - M_x)^2 f_x = 7748,2$; $\sum (x/y - M_x)^2 f_y = 502,99$.

4. По формуле (2.35):

$$\eta_{x/y} = \sqrt{\frac{502,99}{7748,2}} = 0,25.$$

Таблица 2.10. К вычислению обратного корреляционного отношения для взвешенных рядов

	f_x	$x f_x$	x/y	f_y	$(x - M_x)^2 f_x$	$(x/y - M_x)^2 f_y$
1	2	3	4	5	6	7
60	2	120	76,8	14	455,66	40,746
65	29	1885	75,4	38	2954,8	3,5582
70	34	2380	76,3	71	882,26	103,27
75	63	4725	74,6	58	0,55667	14,154
80	74	5920	73,3	26	1781,1	83,679
85	8	680	70,0	6	785,03	155,69
90	4	360	65,0	1	888,76	101,89
	214	16 070	—	214	7748,2	502,99

5. Ошибка отношения и критерий Фишера (2.29), (2.30):

$$m_{\eta^2} = (1 - 0,25^2) (6/207) = 0,027; \quad v_1 = 6; \quad v_2 = 207.$$

$$F = 0,25^2/0,027 = 2,30 > F' = 2,14 \text{ (табл. 9П).}$$

6. Критерий достоверности (2.31):

$$\frac{[(207 - 2) 0,25^2]/[(1 - 0,25^2) 6] - 1}{\sqrt{[2(6 + 207 - 2)]/[6(207 - 4)]}} = 2,17 < 3.$$

Оба критерия позволяют заключить, что обратная зависимость — продолжительность вегетационного периода от веса 1000 семян у ячменя является еще более слабой, чем прямая, поэтому считаем, что аргумент в § 2.10 был назначен верно и что вес семян у ячменя больше зависит от продолжительности их созревания, чем наоборот.

§ 2.12. Полихорический коэффициент сопряженности Чупрова

При измерении силы корреляционной связи между двумя признаками, каждый из которых варьирует по качественным показателям, вычисляется полихорический коэффициент сопряженности, предложенный А. А. Чупровым. Число классов варьирования обоих или одного из коррелирующих признаков при этом должно быть больше двух. Если число классов варьирования по два в каждом ряду, рекомендуется вычислять тетрахорический (четырепольный) вариант этого же коэффициента (см. § 2.13). Коэффициент сопряженности можно вычислить и тогда, когда оба ряда имеют многоклассовое распределение со взвешенными вариантами (по примеру корреляционной решетки в табл. 2.01), причем варианты в рядах могут быть и ранжированными.

Полихорический коэффициент сопряженности вычисляется по формуле:

$$T = \frac{J - 1}{\sqrt{(k_x - 1)(k_y - 1)}}, \quad (2.36)$$

где T — полихорический коэффициент; J — сумма частных, вычисленная в табл. 2.11; k_x — число классов в ряду x ; k_y — число классов в ряду y .

Таблица 2.11. Вычисление суммы частных J для полихорического коэффициента сопряженности

Вес 1000 зерен, y	Вегетационный период, дни, x			n_y
	Скороспелые, до 73 дней	Среднеспелые, 74—84 дня	Позднеспелые, более 85 дней	
Крупнозерные более 50,6 г	2 (4) 0,14	12 (144) 5,14	14 (196) 7,0	28
Среднезерные от 38,0 до 50,6 г	27 (729) 4,93	74 (5476) 37,0	47 (2209) 14,93	148
Мелкозерные менее 37,9 г	9 (81) 2,13	25 (625) 16,45	4 (16) 0,42	38
	38	111	65	214 = N
$\sum (f^2/n_y)$	7,20	58,59	22,35	$k_x = 3$ $k_y = 3$
$\frac{\sum (f^2/n_y)}{n_x}$	0,19	0,53	0,34	$J = 1,06$

Вычисления ведем в следующем порядке.

1. Значения всех вариантов разносим по клеткам корреляционной решетки (табл. 2.11).

2. Суммируем частоты по строкам таблицы, получая n_y , и по столбцам, получая n_x . Сумма n_x должна равняться сумме n_y и является объемом выборки: $N = 214$.

3. Частоты всех клеток таблицы возводим в квадрат и произведения записываем рядом в скобках.

4. Квадраты частот делим на сумму n_y данной строки, в которой расположены частоты, например, квадраты всех частот первой строки делим на 28, получаем:

$$4/28 = 0,14; \quad 144/28 = 5,14; \quad 196/28 = 7,0.$$

Квадраты всех частот второй строки делим на 148:

$$729/148 = 4,93 \text{ и т. д.}$$

Результаты деления записываем в тех же клетках.

5. Суммируем результаты деления по всем столбцам таблицы, получаем $\sum (f^2/n_y) = 0,14 + 4,93 + 2,13 = 7,20$ и т. д.

6. Суммы частных от деления по столбцам делим на суммы частот по этим же столбцам:

$$\frac{\sum (f^2/n_y)}{n_x} = \frac{7,20}{38} = 0,19; \quad \frac{58,59}{111} = 0, \quad \text{и т. д.}$$

7. Суммируем результаты последнего деления по строке, получаем $J = 0,19 + 0,53 + 0,34 = 1,06$.

8. Применяем формулу (2.36):

$$T = \frac{1,06 - 1}{\sqrt{(3-1)(3-1)}} = 0,03.$$

9. Оценку достоверности коэффициента сопряженности производим по критерию хи-квадрат, величина которого равна

$$\chi^2 = N (J - 1), \quad (2.37)$$

где χ^2 — критерий хи-квадрат; N — объем выборки; J — сумма, полученная в итоге табл. (2.11).

Число степеней свободы равно:

$$v = (k_x - 1) (k_y - 1) = (3-1) (3-1) = 4.$$

В рассматриваемом примере по формуле (2.37):

$$\chi^2 = 214 (1,06 - 1,00) = 12,84,$$

что выше значений хи-квадрат первых двух уровней достоверности, которые по табл. 10П при $v = 4$ равны соответственно 9,488 и 13,277.

Следовательно, на вес 1000 зерен у ячменя в небольшой степени, но достоверно влияет продолжительность их созревания.

§ 2.13. Тетрахорический коэффициент сопряженности

Силу связи двух признаков, варьирующих альтернативно, можно определить также при помощи коэффициента сопряженности Чупрова (см. § 2.12), формула которого для четырехпольной таблицы данных следующая:

$$T_4 = \frac{|ad - bc| - 0,5N}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \quad (2.38)$$

где T_4 — тетрахорический коэффициент сопряженности; a — ячейка табл. 2.12, в которую заносятся особи или предметы, обладающие обоими признаками; b — ячейка, в которую заносятся особи, обладающие признаками x и без признака y ; c — ячейка для особей без признака x , но с признаком y ; d — ячейка для особей без обоих признаков; N — объем выборки.

Под x , как и во всех подобных случаях, здесь подразумевается аргумент или причина изучаемого явления, а y — функция, или следствие, условно или фактически. Ряд x располагается горизонтально,

Таблица 2.12. Четырехпольная корреляционная решетка данных к вычислению тетрахорического коэффициента сопряженности

y	Ости имеются	Ости отсутствуют	Σ
Сорта, иммунные к бурой ржавчине	$a = 166$	$c = 34$	200
Сорта, не иммунные к бурой ржавчине	$b = 1297$	$d = 1078$	$b + d$
	$a + b = 1463$	$c + d = 1112$	$N = 2575$

а ряд y — вертикально. Выражение $|ad - bc|$ означает, что разница в прямых скобках в любом случае получает знак плюс. Буквы $a - d$ обозначают также численности частот в этих же ячейках.

Рассмотрим последовательность вычислений на примере измерения силы связи между наличием остей и иммунитетом к бурой ржавчине у сортов пшеницы (табл. 2.12).

1. Располагая классы варьирования, как показано в табл. 2.12, производим разnosку вариантов по ячейкам $a - d$ согласно их назначению, указанному в объяснении к формуле (2.38).

2. Подставляя значения полученных частот в формулу (2.38), получим:

$$T_4 = \frac{|166 \cdot 1078 - 1297 \cdot 34| - 0,5 \cdot 2575}{\sqrt{(166 + 1297)(34 + 1078)(166 + 34)(1297 + 1078)}} = +0,15.$$

3. Оценку значимости полученного коэффициента производим по формуле

$$\chi^2 = NT_4^2, \quad (2.39)$$

где χ^2 — критерий хи-квадрат, используемый здесь для оценки силы сопряженности наблюдаемых факторов; N — объем выборки; T_4 — тетрахорический коэффициент сопряженности.

В данном примере

$$\chi^2 = 2575 \cdot 0,15^2 = 57,9.$$

При оценке значимости коэффициента сопряженности число степеней свободы всегда равно $\nu = 1$. По табл. 10П при доверительном уровне $P_1 = 0,95$, $\chi^2 = 3,841$; при $P_2' = 0,99$, $\chi^2 = 6,635$ и при $P_3' = 0,999$, $\chi^2 = 10,827$. В данном случае вычисленное значение хи-квадрат, равное 57,9, значительно больше всех трех критических значений. Следовательно, можно утверждать, что имеется положительная зависимость иммунитета сортов пшеницы от наличия остей у них. Сила этой связи измеряется тетрахорическим коэффициентом сопряженности связи: $T_4 = +0,15$.

Направление связи можно определить, вычислив ожидаемую частоту ячейки a . Она равна: $a' = \frac{(a+c)(a+b)}{N} = \frac{200 \cdot 1463}{2575} = 113,6$. Так как фактическая частота (166) больше ожидаемой (113,6), то связь положительная. Когда фактическая частота a меньше ожидаемой a' , то связь между признаками отрицательная.

При многозначных вариантах, как это имеет место в рассмотренном примере, формулой (2.38) иногда удобнее пользоваться в логарифмированном виде: $T_4 = \exp 0,5 \{2 \ln (|ad - bc| - 0,5N) - [\ln (a + b) + \ln (c + d) + \ln (a + c) + \ln (b + d)]\}$, где обозначения те же, что и к формуле (2.38).

§ 2.14. Бисериальный коэффициент корреляции

Применяется бисериальный коэффициент корреляции тогда, когда один ряд варьирует по числу классов больше двух, а второй ряд имеет альтернативное (двухклассовое), бисериальное размещение вариантов. В биологии нередки подобные случаи, когда один признак варьирует количественно, а другой, связанный с ним, изменяется только качественно: по принципу «да» или «нет». Коэффициент корреляции по формуле (2.40) можно вычислять для любых взвешенных рядов и при любой численности выборки, в том числе при $N < 100$ (см. начало § 2.15),

$$r_b = \frac{\sum f a_x a_y - (\sum f a_x \sum f a_y) / N}{\sqrt{(\sum f a_x^2 - (\sum f a_x)^2 / N)(\sum f a_y^2 - (\sum f a_y)^2 / N)}}, \quad (2.40)$$

где r_b — бисериальный коэффициент корреляции; f — частоты; a_x — условные отклонения ряда x ; a_y — условные отклонения ряда y ; N — объем выборки.

Вычисления сумм к формуле (2.40) приведены в табл. 2.13. Порядок вычислений следующий.

Таблица 2.13. К вычислению бисериального коэффициента корреляции

y	Даты зацветания, в днях от 1 марта, x							f_y	a_y	$f_y a_y$	$f_y a_y^2$
	57	85	113	141	169	197	225				
Открытый участок	71	172	131	57	13	7	2	459	0	0	0
Теневой участок	71	167	140	63	11	6	1	459	1	459	459
f_x	148	339	271	120	24	13	3	918		459	459
a_x	—3	—2	—1	0	+1	+2	+3	0			
$f_x a_x$	—444	—678	—271	0	24	26	9	—1393	+59	—1334	
$f_x a_x^2$	1332	1356	271	0	24	52	27	3062			

1. Составляем корреляционную решетку, в которой ряд x представляет собой даты зацветания в днях от 1 марта, а y — ряд с альтернативным распределением вариантов на два участка — теневой и освещенный.

2. Суммируем частоты по столбцам и строкам, итоги которых должны совпадать и составляют объем выборки: $N = 918$.

3. Образует условные отклонения a_y , они всегда будут 0 и 1. Отклонения a_x получаются путем поочередного вычитания из всех вариантов ряда x числа $A = 141$, т. е. центральной варианты и деления разностей на классовой интервал, который в данном случае равен: $c = 28$.

4. Перемножив условные отклонения a_x на соответствующие им суммы частот f_x , получим: $-3 \cdot 148 = -444$; $-2 \cdot 339 = -678$; $-1 \cdot 271 = -271$ и т. д.

5. Перемножим квадраты условных отклонений a_x^2 на те же частоты: $-3^2 \cdot 148 = 1332$; $-2^2 \cdot 339 = 1356$ и т. д. Те же самые действия произведем с частотами и условными отклонениями f_y, a_y .

6. Суммируем: $\Sigma f_x a_x = -1334$; $\Sigma f_x a_x^2 = 3062$; $\Sigma f_y a_y = 459$; $\Sigma f_y a_y^2 = 459$.

7. Получаем сумму: $\Sigma f a_x a_y = 71(-3) \cdot 1 + 167(-2) \cdot 1 + 140(-1) \cdot 1 + 63 \cdot 0 \cdot 1 + 11 \cdot 1 \cdot 1 + 6 \cdot 2 \cdot 1 + 1 \cdot 3 \cdot 1 = -661$ путем перемножения каждой частоты из строки «теневой участок» на соответствующие условные отклонения a_x . Частоты из строки «открытый участок» не множим на отклонения, так как все произведения будут равны 0, ибо в ней $a_y = 0$.

8. По формуле (2.40):

$$r_b = \frac{-661 - [(-1334) \cdot 459] / 918}{\sqrt{(3062 - (-1334)^2 / 918)(459 - 459^2 / 918)}} = -0,012.$$

9. Ошибка бисериального коэффициента корреляции ввиду того, что $N > 100$, определяется по формуле (2.17): $m_r = \frac{1 - r^2}{\sqrt{N}} = \frac{1 - 0,012^2}{\sqrt{918}} = 0,033$. Оценка достоверности r_b производится по критерию Стьюдента:

$$t = 0,012 / 0,033 = 0,36.$$

Величины ошибки и критерия Стьюдента совпадают с таковыми, вычисленными по формуле (2.14), с преобразованием через z . По табл. 3П при $\nu = N - 2 = 918 - 2 = 916$ и $P_1 = 95\%$ $t = 1,96$. Поскольку вычисленное значение t меньше табличного, делаем вывод об отсутствии корреляционной связи между числами видов, зацветающих одновременно на теневом и открытом участках. Следовательно, динамика зацветания видов на этих участках различна, на нее больше влияют местные микроусловия, чем общий ход изменения метеофакторов.

§ 2.15. Бисериальный коэффициент корреляции по таблице Келли—Вуда

Этот коэффициент вычисляется в тех же случаях, что и предыдущий (см. § 2.14), но по технике вычислений он проще благодаря наличию специальных таблиц. При $N < 100$ вместо метода из § 2.15 для вычисления бисериального коэффициента лучше пользоваться методом из § 2.14 ввиду довольно сложной поправки, которая вводится в методе, рассматриваемом в § 2.15 при объеме выборки меньше 100. Бисериальный коэффициент корреляции с помощью таблицы находится по формуле:

$$r_b = \frac{|M^- - M^+| \cdot pq}{\sigma_{\Sigma} \xi}, \quad (2.41)$$

где r_b — бисериальный коэффициент корреляции; M^- — взвешенная средняя арифметическая ряда отрицательных частот; M^+ — средняя арифметическая ряда положительных частот; p — доля ряда положительных частот; q — доля ряда отрицательных частот; σ_{Σ} — среднее квадратическое отклонение суммарного ряда частот (ряда Σf в табл. 2.14); ξ — ордината в месте границы классов положительных и отрицательных частот; x — абсцисса этой же точки в долях сигмы.

Таблица 2.14. К вычислению бисериального коэффициента корреляции с использованием таблицы Келли—Вуда

Повреждаемость растений, баллы, x	Условия освещенности растений, y		Σf	Σfx	Σfx^2
	в тени, f_+	на открытом месте, f_-			
3	—	1	1	3	9
4	—	4	4	16	64
5	—	7	7	35	175
6	2	21	23	138	828
7	5	16	21	147	1029
8	2	3	5	40	320
9	—	1	1	9	81
Σ	9	53	62	388	2506

$$M^+ = 7,00 \quad M^- = 6,13$$

ξ , x в соответствии с p находим по таблице Келли—Вуда (табл. 11П). Прямые скобки в формуле (2.41) обозначают, что знак разности средних арифметических не принимается во внимание, т. е. бисериальный коэффициент всегда положительная величина.

Вычисление бисериального коэффициента корреляции ведем в следующем порядке.

1. Заполняем корреляционную решетку по образцу табл. 2.14. В столбец f_+ заносим особи, обладающие первым изучаемым признаком (растут в тени), а в столбец f_- заносим особи со вторым альтерна-

тивным признаком (растут на открытом месте). Требуется выяснить, связана ли повреждаемость растений болезнями с их освещенностью.

2. Суммируем частоты по столбцам, получая суммы $\Sigma f_+ = 9$; $\Sigma f_- = 53$ и по строкам $\Sigma f = N = 62$.

3. Находим величины p и q . Так как следует выполнить условие $p > q$, то для определения q из сумм рядов f_+ и f_- выбираем меньшую, т. е. $\Sigma f_+ = 9$, откуда одна доля равна:

$$q = \frac{\Sigma f_+}{N} = \frac{9}{62} = 0,1452$$

и другая:

$$p = 1 - q = 1 - 0,1452 = 0,8548.$$

4. Определяем средние арифметические рядов f_+ и f_- по формуле (1.13):

$$M^+ = \frac{6 \cdot 2 + 7 \cdot 5 + 8 \cdot 2}{9} = 7,00 \text{ баллов,}$$

$$M^- = \frac{3 \cdot 1 + 4 \cdot 4 + 5 \cdot 7 + 6 \cdot 21 + 7 \cdot 16 + 8 \cdot 3 + 9 \cdot 1}{53} = 6,13 \text{ балла.}$$

Таким образом, повреждаемость растений в тени больше, чем на открытом месте.

5. Вычисляем сигму суммарного ряда частот по формуле (1.38):

$$\sigma_\Sigma = \sqrt{\frac{2506 - \frac{388^2}{62}}{62 - 1}} = 1,13 \text{ балла.}$$

6. По табл. 11П для $p = 0,85$ находим величины $z = 0,2332$ и $x = 1,0364$.

7 По формуле (2.41) бисерпальный коэффициент корреляции

$$r_b = \frac{|6,13 - 7,00|}{1,13 \cdot 0,2332} = 0,41.$$

8. Так как в рассматриваемом случае число наблюдений равно 62 (меньше 100), то к полученному значению коэффициента вычисляется поправка (которая не определяется при $N > 100$) по формуле

$$r_{bc} = r_b \left\{ 1 + \frac{1}{N} \left[0,25 + \frac{pq}{2z^2} - \left(1 - \frac{px}{z} \right) \left(1 + \frac{qx}{z} \right) + 0,25r_b^2 \right] \right\}, \quad (2.42)$$

где r_{bc} — исправленный бисерпальный коэффициент корреляции с учетом числа наблюдений; r_b — коэффициент корреляции до введения поправки; N — объем выборки; остальные обозначения те же, что и к формуле (2.41). Для рассматриваемого примера:

$$r_{bc} = 0,41 \left\{ 1 + \frac{1}{62} \left[0,25 + \frac{0,855 \cdot 0,145}{2 \cdot 0,2332^2} - \left(1 - \frac{0,855 \cdot 1,0364}{0,2332} \right) \times \left(1 + \frac{0,145 \cdot 1,0364}{0,2332} \right) + 0,5 \cdot 0,41^2 \right] \right\} = 0,4502.$$

9. Ошибку бисериального коэффициента корреляции находим по формуле

$$m_r = \frac{(V \overline{pq/3}) - r_b^2}{\sqrt{N}}, \quad (2.43)$$

где обозначения те же, что и к формуле (2.42).

В данном примере

$$m_r = \frac{\sqrt{\frac{0,8548 \cdot 0,1452}{0,2332}} - 0,4502^2}{\sqrt{62}} = 0,1661.$$

10. Ошибку бисериального коэффициента корреляции можно определить также по таблице Сопера (табл. 12П) по формуле

$$m_r = \sqrt{(\lambda_2^2 - \lambda_3 r_b^2 + r_b^4)/N} \quad (2.44)$$

где m_r — ошибка бисериального коэффициента корреляции; λ_2^2 , λ_3 — величины, находимые по табл. 12П в зависимости от величины q ; r_b — величина бисериального коэффициента корреляции; N — объем выборки.

Отсюда ошибка по Соперу равна:

$$m_r = \sqrt{(2,3453 - 3,0923 \cdot 0,4502^2 + 0,4502^4)/62} = 0,1685,$$

т. е. по величине ошибка практически совпадает с вычисленной по формуле (2.43).

11. Если доля q окажется меньше 0,05, то ошибку r_b вычисляют по формуле:

$$m_r = \sqrt{\frac{1}{N} \left\{ \frac{pq}{3^2} - \left[\frac{3}{2} + \left(1 - \frac{px}{3} \right) \left(1 + \frac{qx}{3} \right) \right] r_b^2 + r_b^4 \right\}} \quad (2.45)$$

где обозначения те же, что и к формуле (2.42).

12. Оценка достоверности коэффициента корреляции по критерию Стьюдента:

$$t = r_{bc}/m_r = 0,4502/0,166 = 2,71.$$

При числе степеней свободы: $\nu = N - 2 = 62 - 2 = 60$ и при 95% уровне достоверности по табл. 3П $t = 2,000$. Вычисленное значение больше табличного, следовательно, между степенью затенения растений и степенью их повреждаемости болезнями имеется небольшая, но достоверная связь: $r_b = 0,4502 \pm 0,166$, т. е. чем больше затенение, тем больше повреждаемость растений.

§ 2.16. Частные и множественные коэффициенты корреляции

Основные задачи, которые решают обычно при анализе множественных взаимосвязей — это выявление полной структуры связей в данном комплексе признаков, а также отбор центральных или доминантных, наиболее тесно связанных между собой признаков, служащих

ядром данной структуры или плеяды. Если зависимости между признаками комплекса линейные, что можно предварительно проверить построением простейших графиков, то для исследования структуры взаимосвязей применяют частные и множественные коэффициенты корреляции. Если эти зависимости нелинейные, то их следует привести путем различных преобразований данных к линейной форме. Множественные и частные коэффициенты корреляции возможно вычислить лишь при числе коррелирующих признаков больше двух. Силу совокупной взаимосвязи трех признаков, когда измеряется совместное взаимодействие двух признаков с третьим, можно определить при помощи множественных коэффициентов корреляции по формулам:

$$R_{y \cdot xz}^2 = (r_{xy}^2 + r_{yz}^2 - A)/(1 - r_{xz}^2), \quad (2.46)$$

$$R_{x \cdot yz}^2 = (r_{xy}^2 + r_{xz}^2 - A)/(1 - r_{yz}^2), \quad (2.47)$$

$$R_{z \cdot xy}^2 = (r_{zx}^2 + r_{zy}^2 - A)/(1 - r_{xy}^2). \quad (2.48)$$

Частные коэффициенты корреляции вычисляются по формулам:

$$r_{yx \cdot z} = (r_{yx} - r_{yz}r_{xz})/\sqrt{(1 - r_{yz}^2)(1 - r_{xz}^2)}, \quad (2.49)$$

$$r_{yz \cdot x} = (r_{yz} - r_{yx}r_{xz})/\sqrt{(1 - r_{yx}^2)(1 - r_{xz}^2)}, \quad (2.50)$$

$$r_{xz \cdot y} = (r_{xz} - r_{xy}r_{zy})/\sqrt{(1 - r_{xy}^2)(1 - r_{zy}^2)}, \quad (2.51)$$

где $R_{y \cdot xz}$ — множественный коэффициент корреляции; точка, например, после y означает, что изучается совместное или объединенное влияние аргументов x, z на функцию y ; $A = 2r_{xy}r_{xz}r_{yz}$; r_{yz} , r_{yx} , r_{xz} — полные (обычные) коэффициенты корреляции попарно между признаками x, y, z ; $r_{yz \cdot x}$, $r_{yx \cdot z}$, $r_{xz \cdot y}$ — частные коэффициенты корреляции попарно при исключении влияния третьего признака, индекс последнего ставится справа от точки.

Полные коэффициенты корреляции вычисляются по формуле (2.15), если ряды взвешенные, или по формулам (2.13), (2.18), если ряды невзвешенные.

Вычисление множественного и частных коэффициентов корреляции производится в следующем порядке.

1. Вычисляем полные коэффициенты корреляции между признаками сои, для чего произведем действия, указанные в табл. 2.15. Подставляя суммы столбцов этой таблицы в формулу (2.13), получим

$$\begin{aligned} r_{yx} &= \frac{\sum yx - (\sum y \cdot \sum x)/N}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{N}\right]\left[\sum y^2 - \frac{(\sum y)^2}{N}\right]}} = \\ &= \frac{185412 - (2868 \cdot 839)/13}{\sqrt{\left(55869 - \frac{839^2}{13}\right)\left(634906 - \frac{2868^2}{13}\right)}} = 0,163; \end{aligned}$$

Таблица 2.15. Вычисление полных (обычных) коэффициентов корреляции между тремя признаками

Высота растений, см, x	Вес 1000 зерен, г, y	Урожайность, ц/га, z		y^2		xy		yz
62	217	19	3 844	47 089	361	13 454	1 178	4 123
63	212	20	3 969	44 944	400	13 356	1 260	4 240
73	218	12	5 329	47 524	144	15 914	876	2 616
82	242	17	6 724	58 564	289	19 844	1 394	4 114
66	199	16	4 356	39 601	256	13 134	1 056	3 184
44	200	17	1 936	40 000	289	8 800	748	3 400
43	235	25	1 849	55 225	625	10 105	1 075	5 875
54	235	25	2 916	55 225	625	12 690	1 350	5 875
63	225	19	3 969	50 625	361	14 175	1 197	4 275
74	234	25	5 476	54 756	625	17 316	1 850	5 850
74	223	27	5 476	49 729	729	16 502	1 998	6 021
64	218	23	4 096	47 524	529	13 952	1 472	5 014
77	210	15	5 929	44 100	225	16 170	1 155	3 150
839	2868	260	55 869	634 906	5458	185 412	16 609	57 737

$$r_{xz} = \frac{\sum xz - (\sum x \sum z)/N}{\sqrt{[\sum x^2 - (\sum x)^2/N][\sum z^2 - (\sum z)^2/N]}} =$$

$$= \frac{16609 - (839 \cdot 260)/13}{\sqrt{(55869 - 839^2/13)(5458 - 260^2/13)}} = -0,257;$$

$$r_{yz} = \frac{\sum yz - (\sum y \sum z)/N}{\sqrt{[\sum y^2 - (\sum y)^2/N][\sum z^2 - (\sum z)^2/N]}} =$$

$$= \frac{57737 - (2868 \cdot 260)/13}{\sqrt{\left(5458 - \frac{260^2}{13}\right)\left(634906 - \frac{2868^2}{13}\right)}} = 0,502.$$

Ошибки по формуле (2.14) и значения критерия достоверности с преобразованием через z (табл. 26П) этих коэффициентов следующие:

$$r_{yx} = 0,163; \quad m_z = 1/\sqrt{13-3} = 0,316; \quad t_z = 0,164/0,316 = 0,52.$$

$$r_{yz} = 0,502; \quad m_z = 0,316; \quad t_z = 0,552/0,316 = 1,75.$$

$$r_{xz} = -0,257; \quad m_z = 0,316; \quad t_z = 0,263/0,316 = 0,83.$$

2. Все три полных коэффициента корреляции оказались недостоверными, так как величины t_z меньше табличных, однако это не означает, что связь между рассматриваемыми признаками отсутствует. Например, связь между y и z , т. е. весом 1000 семян и урожайностью, хотя и недостоверна, но сильнее, чем прочие связи. Может возникнуть вопрос: не мешает ли полнее выявить связи между y и z влияние x , т. е. высоты растений. Чтобы проверить эту гипотезу, необходимо исключить из связи между y и z влияние x . Для этого

вычисляем частные коэффициенты корреляции, исключая поочередно влияние каждого фактора из трех по формулам (2.49)—(2.51):

$$r_{yx \cdot z} = \frac{0,163 - 0,502 \cdot (-0,257)}{\sqrt{(1 - 0,502^2)[1 - (-0,257)^2]}} = 0,350;$$

$$r_{yz \cdot x} = \frac{0,502 - 0,163 \cdot (-0,257)}{\sqrt{(1 - 0,163^2)[1 - (-0,257)^2]}} = 0,571;$$

$$r_{xz \cdot y} = \frac{-0,257 - 0,163 \cdot 0,502}{\sqrt{(1 - 0,163^2)(1 - 0,502^2)}} = -0,404.$$

Исключаемый фактор в индексах при r обозначен буквой справа от точки. Критерий достоверности (t) частного коэффициента корреляции при любом объеме совокупности вычисляют по формуле:

$$t = r_{xy \cdot z} \sqrt{(N - n - 2)/(1 - r_{xy \cdot z}^2)}, \quad (2.52)$$

подставляя в которую известные величины, получим:

$$t_{xy \cdot z} = 0,35 \sqrt{(13 - 1 - 2)/(1 - 0,35^2)} = 1,182;$$

$$t_{yz \cdot x} = 0,571 \sqrt{(13 - 1 - 2)/(1 - 0,571^2)} = 2,200;$$

$$t_{xz \cdot y} = -0,404 \sqrt{(13 - 1 - 2)/(1 - 0,571^2)} = -1,400.$$

В формулу (2.52) вместо $r_{xy \cdot z}$ можно, конечно, подставлять величину любого частного коэффициента корреляции.

При оценке достоверности частного коэффициента корреляции по табл. 3П число степеней свободы принимается: $\nu = N - 2 - n$, где N — объем выборки; n — число факторов, действие которых устраняется, в данном случае $n = 1$, следовательно: $\nu = 13 - 2 - 1 = 10$; то же значение буквы N , n имеют в формуле (2.52).

Таким образом, поочередное исключение влияния отдельных факторов повышает силу связи между двумя остальными, хотя эта связь и осталась недоказанной (по-видимому, из-за недостаточного объема выборки), кроме связи между y и z , которая действительно стала достоверной, когда мы устранили влияние высоты растений (x) на связь между весом 1000 семян (y) и урожайностью (z). Следовательно, частные коэффициенты корреляции имеют важное самостоятельное значение как метод исследования множественных корреляций.

3. Множественные коэффициенты корреляции, отражающие силу взаимосвязи одного с двумя остальными факторами совокупно, определим по формулам (2.46)—(2.48):

$$R_{y \cdot xz}^2 = (0,163^2 + 0,502^2 - A)/[1 - (-0,257)^2] = 0,34330;$$

$$R_{y \cdot xz} = 0,58593;$$

$$R_{x \cdot yz}^2 = [0,163^2 + (-0,257)^2 - A]/(1 - 0,502^2) = 0,18005;$$

$$R_{x \cdot yz} = 0,42432;$$

$$R_{z \cdot xy}^2 = (-0,257^2 + 0,502^2 - A)/(1 - 0,163^2) = 0,36994;$$

$$R_{z \cdot xy} = 0,60822;$$

$$A = 2 \cdot 0,163 \cdot 0,502 \cdot (-0,257) = -0,042058.$$

Без применения таблиц достоверность коэффициента множественной корреляции оценивается по критерию Романовского:

$$\frac{(N-5) R^2 [2(1-R^2)]^{-1} - 1}{\sqrt{(N-3)/(N-7)}} \geq 3, \quad (2.53)$$

где N — объем выборки; R — коэффициент множественной корреляции. Из формулы (2.53) видно, что коэффициент множественной корреляции может быть достоверным лишь в том случае, если объем совокупности не менее 8. Если левая часть формулы (2.53) будет больше или равна 3, то коэффициент считается значимым. Величина $R_{y \cdot xz}$ недостоверна, так как:

$$\frac{(13-5) 0,586^2 \cdot [2(1-0,586^2)]^{-1} - 1}{\sqrt{(13-3)/(13-7)}} = 0,84 < 3.$$

Величина этого критерия для $R_{x \cdot yz}$ равна $-0,094$, для $R_{z \cdot xy}$ она составляет наибольшую величину: $1,04$. Следовательно, для уравнения множественной регрессии этих признаков z следует принять за функцию, т. е. искать зависимость: $z' = a_0 + a_1x + a_2y$. При оценке достоверности величины множественного коэффициента корреляции можно пользоваться также критерием Фишера по формуле:

$$F = \frac{R^2}{1-R^2} \cdot \frac{N-n}{n-1}, \quad (2.54)$$

где F — вычисляемая величина критерия, которая сравнивается с табличной (табл. 9П) при $\nu_1 = n - 1$, $\nu_2 = N - n$, степенях свободы; R — множественный коэффициент корреляции; N — объем совокупности; n — число признаков в изучаемом комплексе.

Для $R_{x \cdot yz} = 0,42432$

$$F_x = \frac{0,42432^2}{1-0,42432^2} \cdot \frac{13-3}{3-1} = 1,10;$$

$$\text{для } R_{y \cdot xz} = 0,58593 \quad F_y = 2,61;$$

$$\text{для } R_{z \cdot xy} = 0,60822 \quad F_z = 2,94.$$

По табл. 9П при $\nu_1 = 3 - 1 = 2$ и $\nu_2 = 13 - 3 = 10$ на 95% доверительном уровне $F' = 4,10$, что больше каждой из трех вычисленных величин критерия. Следовательно, все три найденных множественных коэффициента корреляции недостоверны при оценке как критерием Романовского, так и критерием Фишера. Причина недостоверности связи в данном случае, по-видимому, заключается в недостаточном объеме совокупности.

Таким образом, все нужные величины для построения структуры взаимосвязей в изучаемом комплексе вычислены — это частные и множественные коэффициенты корреляции, которые отражают как попарные связи признаков, так и связи совместно двух из них с третьим.

РЕГРЕССИЯ

После того как при помощи методов из главы 2 установлено наличие корреляционной связи между двумя изучаемыми явлениями или признаками, можно попытаться установить закономерность количественного изменения одного из признаков (функция) при изменении другого (аргумент). Эту закономерность можно далее использовать для прогноза изучаемого явления или выяснения его критических точек. С этой целью изучаемую связь выражают аналитически — в виде соответствующего уравнения регрессии и графически — с вычислением точек вычерчиваемой теоретической кривой по найденному уравнению. В этой главе даются методы такого аналитического и графического выражения некоторых типов корреляционных зависимостей.

Определенного ответа на вопрос о достаточном числе точек для построения той или иной кривой регрессии дать нельзя, так как это связано с числом коэффициентов уравнения и точек перегиба кривой, с одной стороны, и дисперсией исходных данных, с другой стороны. Чем больше число перегибов у кривой и больше дисперсия данных, тем больше требуется точек для вычисления уравнения и графического представления изучаемой зависимости. Поэтому, хотя формально прямую линию можно построить и по двум точкам, а параболу второго порядка — по трем, на практике, особенно в биологических исследованиях, почти никогда не бывает достаточно этого минимума точек, чтобы более или менее ясно представить себе тенденцию изучаемой связи. В общем случае рассматриваемый вопрос следует решать в начале исследования, когда определяется репрезентативность выборок. Если объем обеих выборок достаточен для достоверности определения их основных статистических показателей — средней и сигмы, то объемы этих выборок обычно достаточны и для выявления тенденции взаимосвязи между ними.

Независимо от способа расчета коэффициентов уравнений регрессии существенно снижает трудоемкость расчетов преобразование исходных данных так, чтобы их величина приближалась по возможности к первому порядку. Для этого исходные данные делят на одно и то же число или вычитают из них среднее или начальное значение, а также принимают другие единицы измерения.

§ 3.01. Уравнение прямой линии

Эмпирические линии регрессии, получаемые по способу из § 2.01, отражают изучаемую корреляционную связь с присущими ей случайными отклонениями и не дают возможности аналитического про-

гноза данного явления. Чтобы получить уравнение, по которому можно такой прогноз сделать, требуется аппроксимировать данную эмпирическую линию регрессии подходящей теоретической линией регрессии, которая даст возможность аналитической интерпретации связи. Простейшей теоретической линией регрессии, подходящей для приблизительного прогноза явления, в некоторых случаях является прямая линия, или парабола первого порядка. Уравнение прямой линии является наиболее простым для расчета, поэтому путем различных преобразований данных часто стремятся свести криволинейные зависимости к прямолинейным. Однако нередко корреляционные связи в биологии имеют такую форму криволинейности, что уравнение прямой к ним применить не удастся. В таких случаях, изучив форму эмпирической линии регрессии (см. § 2.01), следует попытаться подобрать к ней подходящую теоретическую кривую (см. § 3.03 и 3.14).

На примере зависимости веса 1000 семян от продолжительности вегетационного периода у ячменя ниже излагаются пять способов вычисления уравнения прямой линии, которое в общей форме имеет вид:

$$y' = a_0 + a_1x, \quad (3.01)$$

где y' — теоретические значения функции или зависимой переменной; x — аргумент или независимая переменная; a_0, a_1 — коэффициенты уравнения, имеющие различное значение в зависимости от специфики изучаемого явления; y — эмпирические значения зависимой переменной (в табл. 3.01).

Иногда уравнение прямой линии задают в форме: $\frac{x}{a} + \frac{y}{b} = 1$, где a — абсцисса точки пересечения прямой с осью абсцисс; b — ордината точки пересечения прямой с осью ординат. Этой формой, называемой уравнением прямой в отрезках, в биометрии обычно не пользуются.

Таблица 3.01. К вычислению уравнения прямой линии

Продолжительность вегетационного периода, x	Вес 1000 семян ячменя, y		yx	y'	$y' - y$	$(y' - y)^2$
1	2	3	4	5	6	7
90	47,50	8100	4275,00	48,3399	0,84	0,7056
85	46,75	7225	3973,75	46,8004	0,05	0,0025
80	45,75	6400	3660,00	45,2609	-0,49	0,2401
75	42,85	5625	3213,75	43,7214	0,87	0,7569
70	44,76	4900	3133,20	42,1819	-2,58	6,6564
65	41,44	4225	2693,60	40,6424	-0,80	0,6400
60	37,00	3600	2220,00	39,1029	2,10	4,4100
525	306,05	40 075	23 169,30	306,0498	-0,01	13,4115

Для того чтобы можно было пользоваться уравнением (3.01) в применении к вышеназванной конкретной связи, необходимо определить присущие ей величины коэффициентов a_0 и a_1 .

Первый способ. Определение коэффициентов уравнений теоретических линий регрессии производится различными способами: самый точный из них — метод наименьших квадратов, посредством которого вычислим коэффициенты уравнения прямой линии.

1. В качестве исходных данных для нахождения коэффициентов уравнения служат точки эмпирической линии регрессии (способ ее вычисления см. в § 2.01), если функция и аргумент представлены взвешенными рядами. Если же выборки y и x представлены невзвешенными рядами, то для расчета коэффициентов уравнения регрессии берутся непосредственно варианты этих рядов, без вычисления эмпирической линии регрессии.

Требующиеся здесь значения эмпирической линии регрессии y и x уже подсчитаны в табл. 2.01.

2. В столбцах 1 и 2 табл. 3.01 приведены значения ряда x и y/x из табл. 2.01. В столбце 3 и 4 производятся действия, суть которых ясна из заголовков столбцов.

3. Суммы, полученные в итоге столбцов 1—4, подставляем в систему двух нормальных уравнений:

$$(1) \quad a_0 N + a_1 \Sigma x = \Sigma y,$$

$$(2) \quad a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy, \quad (3.02)$$

получаем:

$$(1) \quad 7a_0 + 525a_1 = 306,05;$$

$$(2) \quad 525a_0 + 40075a_1 = 23169,3.$$

4. Находим значения a_0 и a_1 из уравнений следующим образом. Делим число при a_0 во втором уравнении на число при a_0 в первом уравнении: $525 : 7 = 75$; умножаем все члены первого уравнения на 75; вычитаем из второго уравнения первое:

$$\begin{array}{r} 525a_0 + 40075a_1 = 23169,3 \\ - 525a_0 + 39375a_1 = 22953,75 \\ \hline 700a_1 = 215,55, \end{array}$$

откуда $a_1 = 0,3079$. Подставляя найденное значение a_1 в первое уравнение, находим:

$$a_0 = (306,05 - 525 \cdot 0,3079) / 7 = 20,6289.$$

Уравнение (3.01) получает следующие коэффициенты:

$$y' = 20,6289 + 0,3079x \quad \text{или} \quad \text{округленно:} \quad y' = 20,63 + 0,31x.$$

5. В столбце 5 вычислены теоретические значения веса 1000 семян y' при различной продолжительности вегетационного периода x .

6. В столбце 6 приводится разница между теоретическими и фактическими значениями веса 1000 семян.

Ошибка уравнения прямолинейной или криволинейной регрессии вычисляется по формуле:

$$m_{y \cdot x} = \sigma_y \sqrt{\frac{(N-1)(1-\eta_{y/x}^2)}{N-n}}, \quad (3.03)$$

или по вытекающей из нее формуле:

$$m_{y \cdot x} = \sqrt{\frac{\sum (y - y')^2}{N-n}}, \quad (3.04)$$

где $m_{y \cdot x}$ — ошибка уравнения регрессии; y — эмпирические значения функции; y' — теоретические значения функции; N — число

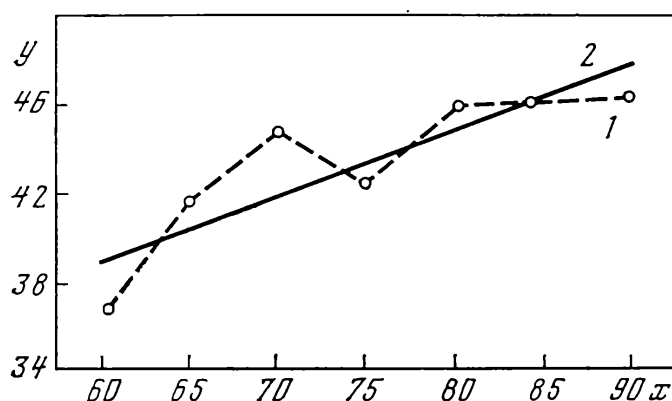


Рис. 16. Зависимость между продолжительностью вегетационного периода (x) и весом 1000 семян (y) у сортов ячменя
1 — эмпирическая линия регрессии y/x ;
2 — прямолинейная регрессия связи

точек эмпирической линии регрессии, по которым вычислялось уравнение регрессии; σ_y — сигма ряда функции (y); $\eta_{y/x}$ — корреляционное отношение, вместо него подставляют r_{xy} при прямолинейной регрессии; n — число коэффициентов уравнения, включая свободный член.

В рассматриваемом примере (табл. 3.01)

$$m_{y \cdot x} = \sqrt{\frac{13,4115}{7-2}} = 1,6 \text{ г.}$$

Таким образом, вес 1000 семян (в г, y) у ячменя можно приблизительно определить, зная продолжительность их созревания (в днях, x) по формуле $y = 20,63 + 0,31x$, с ошибкой $\pm 1,6$ г (рис. 16).

Точки пересечения с осями ординат и абсцисс равны соответственно:

$$y = a_0, \quad x = -(a_0/a_1), \quad (3.05)$$

где обозначения те же, что и к формуле (3.01), т. е. $y = 20,63$; $x = -(20,63/0,31) = -66,5$.

Коэффициент a_1 — основной параметр уравнения прямой линии, степень его достоверности отражает также и наличие или отсутствие корреляционной связи между признаками. Поэтому рекомендуется оценивать достоверность отличия найденного коэффициента a_1 от

нуля по формуле

$$t = \frac{a_1 \sigma_x \sqrt{N-1}}{m_{y \cdot x}}, \quad (3.06)$$

где t — величина критерия Стьюдента, сравниваемая с критической по табл. 3П при числе степеней свободы $\nu = N - 2$; если вычисленная величина меньше табличной, то связь между x , y и значение a_1 достоверны, а если вычисленная будет больше табличной величины, то связь данных признаков и значение a_1 — недостоверны; a_1 — коэффициент при аргументе в уравнении прямой линии; σ_x — среднее квадратическое отклонение ряда аргумента; $m_{y \cdot x}$ — ошибка уравнения по формуле (3.04); N — объем выборки.

Для нашего примера величина $\sigma_x = 10,802$ (см. третий способ в этом параграфе), остальные величины известны, подставив их в формулу (3.06), получим:

$$t = (0,308 \cdot 10,802 \sqrt{7-1})/1,637 = 5.$$

По табл. 3П значение критерия при числе степеней свободы: $\nu = 7 - 2 = 5$ и 99% доверительном уровне меньше вычисленного: $4,032 < 5$. Следовательно, корреляция между x и y действительно существует, и величина коэффициента пропорциональности данной линии регрессии a_1 достоверно отличается от нуля.

Достоверность отличия от нуля коэффициента a_0 можно оценить по формуле

$$t = \frac{a_0}{m_{y \cdot x} \sqrt{\frac{1}{N} + \frac{1}{N-1} \left(\frac{M_x}{\sigma_x} \right)^2}}, \quad (3.07)$$

где t — величина критерия Стьюдента, сравниваемая с табличной (табл. 3П) при числе степеней свободы: $\nu = N - 2$; a_0 — величина свободного члена в уравнении прямой линии; N — объем выборки; M_x , σ_x — средняя арифметическая и сигма ряда x .

$$t = \frac{20,628}{1,637 \sqrt{\frac{1}{7} + \frac{1}{7-1} \left(\frac{75}{10,802} \right)^2}} = 4,4.$$

Полученная величина критерия больше табличной: $4,4 > 4,032$ при 5 степенях свободы и 99% доверительном уровне, поэтому считаем достоверной найденную величину коэффициента a_0 . Для вычисления коэффициентов уравнения прямой линии применяют также следующие способы.

Второй способ. Коэффициенты a_0 и a_1 можно найти по формулам:

$$a_0 = M_y - a_1 M_x, \quad (3.08)$$

$$a_1 = \frac{\sum (x - M_x)(y - M_y)}{\sum (x - M_x)^2}, \quad (3.09)$$

Таблица 3.02. К вычислению коэффициентов уравнения прямой линии по второму способу

	y	$x - M_x$	$(x - M_x)^2$	$y - M_y$	$(x - M_x)(y - M_y)$	$(y - M_y)^2$
90	47,50	15	225	3,777	+56,655	14,266
85	46,75	10	100	3,027	+30,270	9,163
80	45,75	5	25	2,027	+10,135	4,109
75	42,85	0	0	-0,873	0	0,762
70	44,76	-5	25	1,037	-5,185	1,075
65	41,44	-10	100	-2,283	+22,83	5,212
60	37,0	-15	225	-6,723	+100,845	45,199
525	306,05	0	700		+215,55	79,786

где M_y и M_x — средние арифметические рядов y и x . По данным табл. 3.02 $M_x = \Sigma x/N = 525/7 = 75$ дней; $M_y = 306,05/7 = 43,723$ г; $\Sigma (x - M_x)^2 = 700$; $\Sigma (x - M_x)(y - M_y) = 215,55$. $a_1 = 215,55/700 = +0,30793$; $a_0 = 43,723 - 0,30793 \cdot 75 = 20,628$.

Ошибка уравнения и точки пересечения прямой с осями координат вычисляются по тем же формулам (3.04), (3.05).

Третий способ. Если известны средние квадратические отклонения рядов x , y и коэффициент корреляции между ними, то величины коэффициентов a_1 и a_0 уравнения прямой линии, не вычисляя точек эмпирической линии регрессии, можно найти по формулам:

$$a_1 = r_{xy} \frac{\sigma_y}{\sigma_x}, \quad (3.10)$$

$$\text{или } a_1 = r_{xy} \frac{s_y}{s_x} \frac{c_y}{c_x}, \quad (3.11)$$

где a_1 — коэффициент регрессии при аргументе x ; r_{xy} — коэффициент корреляции между признаками x и y ; σ_y , σ_x — сигмы рядов y и x в исходных единицах; s_y , s_x — сигмы тех же рядов в классовых интервалах; c_y , c_x — классовые интервалы рядов y и x .

Свободный член уравнения a_0 вычисляется по формуле (3.08) из второго способа.

Рассчитаем вначале коэффициенты уравнения для той же зависимости веса семян (y) от продолжительности вегетационного периода (x) у ячменя, рассмотренной в первом и втором способах, но с тем отличием, что используем все даты выборки объемом $N = 214$, не вычисляя эмпирической линии регрессии. В § 2.05 для нашего примера уже вычислены: $r_{xy} = 0,2055$; $s_y = 1,203$; $s_x = 1,195$; $c_y = 6$; $c_x = 5$. По формуле (3.11): $a_1 = 0,2055 \frac{1,203}{1,195} \frac{6}{5} = 0,24826$.

Средняя арифметическая вариант вегетационного периода ранее вычислялась в разных примерах (см., например, § 1.10): $M_x = 75,0$. Средняя арифметическая веса семян у ячменя найдена

в § 2.10, она может быть вычислена также по формуле (1.37) с использованием данных из табл. 2.05: $M_y = 40 + (6 \cdot 148)/214 = 44,15$ г, где 148 — сумма второго столбца справа. По формуле (3.08): $a_0 = 44,15 - 0,24826 \cdot 75 = 25,53$. Искомое уравнение прямой линии для $N = 214$: $y = 25,53 + 0,24826x$.

Ошибка уравнения регрессии определяется по формуле

$$m_{y \cdot x} = \sigma_y \sqrt{1 - r^2}, \quad (3.12)$$

где $m_{y \cdot x}$ — ошибка теоретических значений функции; σ_y — среднее квадратическое отклонение ряда y ; r — коэффициент корреляции между x и y .

$$m_{y \cdot x} = 7,187 \sqrt{1 - 0,2055^2} = 7,03.$$

Вместо коэффициента корреляции в формулу (3.12) можно подставлять также корреляционное отношение, измеряющее силу связи между теми же признаками при том же объеме выборки (оно вычислено для этих рядов в § 2.10: $\eta_{y/x} = 0,26$). Отсюда по формуле (3.12): $m_{y \cdot x} = 7,187 \cdot \sqrt{1 - 0,26^2} = 6,94$.

Формула (3.12) представляет собой упрощенный вариант (3.03), применяемый для сравнительно больших выборок. Для невзвешенных рядов x и y , относящихся к той же самой выборке и представляющих собой точки эмпирической линии регрессии, вычисленной по методу из § 2.01, по формуле (1.24), используя суммы столбцов 4 и 7 табл. 3.02: $\sigma_x = 10,802$; $\sigma_y = 3,6466$. Коэффициент корреляции вычислим по формуле (2.18):

$$r = \frac{\sum (x - M_x)(y - M_y)}{\sqrt{\sum (x - M_x)^2 \sum (y - M_y)^2}},$$

подставив в которую данные из табл. 3.02, получим:

$$r = \frac{215,55}{\sqrt{700 \cdot 79,786}} = \frac{215,55}{236,33} = + 0,91207,$$

затем по формуле (3.10):

$$a_1 = 0,9121 \cdot \frac{3,6466}{10,802} = 0,308.$$

Если число вариантов в обоих рядах одинаково, как это обычно и бывает при анализе их взаимосвязей, то коэффициент при x в уравнении прямой линии можно вычислить также по формуле:

$$a_1 = r_{xy} \frac{m_{My}}{m_{Mx}}, \quad (3.13)$$

где m_{My} , m_{Mx} — ошибки средних арифметических рядов x и y , вычисленные по формуле (1.09), подставляя значения которых для

рассматриваемого примера в формулу (3.13), получим:

$$a_1 = 0,9121 \frac{1,3783}{4,0828} = 0,308.$$

a_0 по формуле (3.08) равно тому же значению, что во втором способе.

Для оценки меры соответствия изучаемой взаимосвязи закону прямой линии вычисляют показатель линейности:

$$k_1 = \sqrt{N(1 - r^2)} \geq 3. \quad (3.14)$$

Значение k_1 больше трех указывает на то, что связь не может считаться прямолинейной. Подставив в формулу (3.14) известные величины, получим: $k_1 = \sqrt{7(1 - 0,9121)^2} = 1,09$, что меньше 3; следовательно, рассматриваемая связь может быть интерпретирована в заданных пределах уравнением прямой линии. Ошибка ($m_{y \cdot x}$) определения значений y' по этому уравнению равна:

$$m_{y \cdot x} = \sigma_y \sqrt{\frac{(N-1)(1 - r_{xy}^2)}{N-2}}, \quad (3.15)$$

где все величины уже известны из предыдущих вычислений по третьему способу;

$$m_{y \cdot x} = 3,6466 \sqrt{\frac{(7-1)(1 - 0,9121^2)}{7-2}} = 1,637 = 1,6 \text{ г.}$$

Формула (3.15) представляет собой вариант формулы (3.03), предназначенный только для прямолинейной регрессии.

Как видим, первое уравнение, рассчитанное непосредственно по всем данным выборки, в значительной степени отражает случайные отклонения, большей частью мешающие выявлению основной тенденции зависимости, и ошибка прогноза по нему поэтому больше ($m_{y \cdot x} = 7,03$), чем по уравнению, вычисленному для точек эмпирической линии регрессии: $y' = 20,628 + 0,30793x$; $m_{y \cdot x} = 1,6 \text{ г.}$

Далее в этом параграфе будут сравнены эти два уравнения, чтобы оценить правомерность процедуры замены взвешенного ряда эмпирической линией регрессии.

Третий способ удобен тем, что посредством него можно одновременно вычислить величины средних арифметических, дисперсий, средних квадратических отклонений обоих рядов, коэффициента корреляции между ними, коэффициентов уравнения прямой линии и ошибку последнего.

Четвертый способ. Если не предъявляется особых требований к точности определения коэффициентов уравнения прямой линии, то можно их найти непосредственно по графику эмпирических данных. Во многих случаях, особенно при удачном подборе положения глазомерно проводимой прямой, по этому способу можно получить удовлетворительные результаты.

На рис. 17 эмпирическая линия регрессии, представленная в виде точек, отражает зависимость даты конца цветения от даты зацветания в совокупности из 1351 вида и сорта травянистых многолетников в Москве. Эта зависимость очень близка к прямолинейной и глазо-

Таблица 3.03. Прямолинейная зависимость даты конца цветения от даты зацветания в днях от 1 марта у травянистых многолетников

Начало цветения,	Конец цветения, y	y'	y''
40	75,7	73	75,4
60	92,4	93	95,4
80	114,5	113	115,4
100	134,3	133	135,4
120	159,4	153	155,4
140	176,4	173	175,4
160	203,6	193	195,4
180	216,7	213	215,4
200	230,9	233	235,4
220	230		

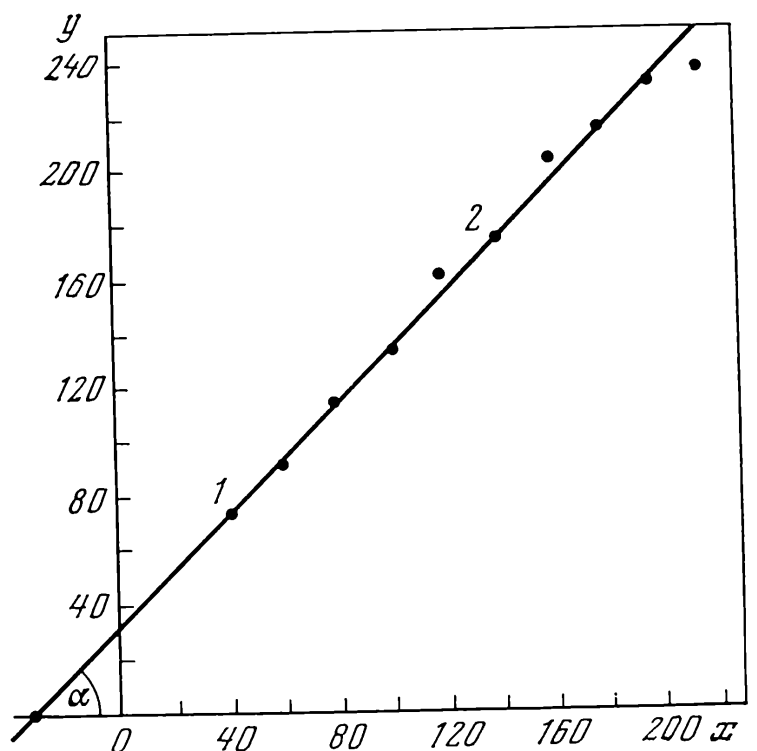


Рис. 17. Зависимость дат (в днях от 1 марта) конца цветения (y) от дат его начала (x) у травянистых многолетников в Москве

мерно выровнена прямой. Измерив угол, под которым эта прямая пересекает ось абсцисс, и определив по таблицам тригонометрических функций тангенс этого угла, тем самым найдем величину коэффициента при x в уравнении. В нашем примере (рис. 17) $\alpha = 45^\circ$, следовательно, $a_1 = \operatorname{tg} \alpha = \operatorname{tg} 45^\circ = 1$. Другой коэффициент равен отрезку от начала координат до места пересечения прямой с осью ординат, т. е. $a_0 = 33$ дня. Таким образом, уравнение данной зависимости: $y' = 33 + x$, где y' — дата конца цветения вида или сорта травянистого многолетника; x — дата его зацветания. В столбце 3 табл. 3.03 вычислены по этому уравнению теоретические значения дат конца цветения, довольно близкие к наблюдаемым срокам.

Пятый способ (избранных точек). На том же примере (табл. 3.03) рассмотрим технику вычислений коэффициентов уравнения прямой линии посредством двух избранных точек (1 и 2 на рис. 17). Координаты этих двух точек: $x_1 = 40$; $y_1 = 75,7$; $x_2 = 140$; $y_2 = 176,4$. Искомые коэффициенты найдем по формулам:

$$a_0 = \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1}, \quad (3.16)$$

$$a_1 = \frac{y_2 - y_1}{x_2 - x_1}, \quad (3.17)$$

т. е.:

$$a_0 = \frac{140 \cdot 75,7 - 40 \cdot 176,4}{140 - 40} = 35,4,$$

$$a_1 = \frac{176,4 - 75,7}{140 - 40} = 1,$$

и уравнение прямой приближенно: $y'' = 35,4 + x$. В четвертом столбце табл. 3.03 приведены даты конца цветения, вычисленные по найденному уравнению, которые несколько лучше подходят к эмпирическим данным, чем полученные по предыдущему способу. Так как уравнения прямой, полученные по способам 4 и 5, являются приближенными, вычисление ошибок прогноза для них нецелесообразно. Чем лучше выражена прямолинейная зависимость у эмпирических данных, тем меньше ошибки определения коэффициентов уравнения прямой линии по обоим последним способам. В § 3.05 приведен алгоритм вычисления коэффициентов уравнений парабол, в том числе прямой линии, по способу сумм.

Сравнение двух уравнений прямолинейной регрессии. При помощи излагаемого метода можно сравнить любые два уравнения прямой линии, чтобы аналитически оценить степень близости линий регрессии, которые они представляют.

Рассмотрим технику расчета критериев для сравнения на примере двух уравнений, полученных разными способами, но по одним и тем же данным. Первое из уравнений: $y = 25,53 + 0,24826x$ получено в этом параграфе по третьему способу прямым путем по всем данным выборки, объем которой $N_1 = 214$. Второе уравнение: $y = 20,628 + 0,30793x$ рассчитано в § 3.01 (первый — третий способы) для той же выборки, но по эмпирической линии регрессии, т. е. фактически лишь по семи точкам, $N_2 = 7$, хотя косвенно и они отражают присущие данной выборке средний уровень варьирования и усредненную тенденцию изменения веса семян (функции) в зависимости от продолжительности вегетационного периода (аргумента) у ячменя. Таким образом, надлежит выяснить, насколько репрезентативно отражает указанную зависимость уравнение, рассчитанное по точкам эмпирической линии регрессии, иначе говоря, можно ли считать это уравнение несущественно отличающимся от того, которое было получено прямым способом, по всем датам этой же выборки. Сравнение названных уравнений проводится в три этапа.

1. Проверка гипотезы о близости величин остаточных дисперсий.

Сравнение остаточных дисперсий, или квадратов ошибок уравнений, производится посредством критерия Фишера по формуле

$$F = \frac{m_{y \cdot x1}^2}{m_{y \cdot x2}^2}, \quad (3.18)$$

где F — величина критерия Фишера; $m_{y \cdot x1}$, $m_{y \cdot x2}$ — ошибки первого и второго уравнений, которые были вычислены раньше и равны соответственно 7,033 и 1,637 г, откуда по формуле (3.18):

$$F = \frac{7,033^2}{1,637^2} = \frac{49,47}{2,68} = 18,5.$$

Объем первой выборки $N_1 = 214$, второй $N_2 = 7$ и число степеней свободы равно: $\nu_1 = N_1 - 2 = 214 - 2 = 212$; $\nu_2 = N_2 - 2 = 7 - 2 = 5$. По табл. 9П при $P_3 = 99,9\%$ значение критерия для указанных чисел свободы равно $F' = 23,78$, т. е.

больше вычисленного ($F = 18,5$). Следовательно, при данной, сравнительно нестрогой оценке остаточные дисперсии двух уравнений или их ошибки различаются несущественно.

2. Гипотеза о близости величин коэффициентов регрессии уравнений проверяется по критерию Стьюдента:

$$|t| = \frac{a_{1.1} - a_{1.2}}{\sigma \sqrt{\frac{1}{(N_1 - 1) \sigma_{x1}^2} + \frac{1}{(N_2 - 1) \sigma_{x2}^2}}}, \quad (3.19)$$

где

$$\sigma = \sqrt{\frac{(N_1 - 2) m_{y \cdot x1}^2 + (N_2 - 2) m_{y \cdot x2}^2}{N_1 + N_2 - 4}}$$

и соответственно для первого и второго уравнений: $a_{1.1}$, $a_{1.2}$ — коэффициенты регрессии, или коэффициенты при x ; σ_{x1}^2 , σ_{x2}^2 — дисперсии аргумента, т. е. величин x ; σ — усредненная величина среднего квадратического отклонения; N_1 , N_2 — объемы выборок; $m_{y \cdot x1}^2$, $m_{y \cdot x2}^2$ — квадраты ошибок; $|t|$ — величина критерия Стьюдента без знака.

$$\sigma = \sqrt{\frac{(214 - 2) 49,47 + (7 - 2) 2,68}{214 + 7 - 4}} = 6,96;$$

$$|t| = \frac{0,24826 - 0,308}{6,96 \sqrt{\frac{1}{(214 - 1) 5,989^2} + \frac{1}{(7 - 1) 10,802^2}}} = 0,217.$$

Число степеней свободы в данном случае равно: $\nu = N_1 + N_2 - 4 = 214 + 7 - 4 = 217$. По табл. ЗП для этого числа степеней свободы и доверительного уровня $P_1 = 95\%$ величина критерия равна 1,960, что больше вычисленной величины t . Следовательно, можем считать две сравниваемые прямые — достаточно параллельными линиями, так как их коэффициенты регрессии несущественно различаются друг от друга.

3. Гипотеза о близости величин свободных членов уравнений (a_{01} и a_{02}) проверяется также по критерию Стьюдента:

$$|t| = \frac{\frac{(N_1 - 1) \sigma_{x1}^2 a_{1.1} + (N_2 - 1) \sigma_{x2}^2 a_{1.2}}{(N_1 - 1) \sigma_{x1}^2 + (N_2 - 1) \sigma_{x2}^2} - \frac{M_{y1} - M_{y2}}{M_{x1} - M_{x2}}}{\sigma \sqrt{\frac{1}{(N_1 - 1) \sigma_{x1}^2 + (N_2 - 1) \sigma_{x2}^2} + \frac{1}{(M_{x1} - M_{x2})^2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}, \quad (3.20)$$

где

$$\sigma = \sqrt{\frac{(N_1 - 2) m_{y \cdot x1}^2 + (N_2 - 2) m_{y \cdot x2}^2}{N_1 + N_2 - 4}}$$

соответственно для первого и второго уравнений: N_1 , N_2 — объемы выборок; σ_{x1}^2 , σ_{x2}^2 — дисперсии аргумента (ряда x); $a_{1.1}$, $a_{1.2}$ —

коэффициенты регрессии; M_{y1}, M_{y2} — средние арифметические функции (ряда y); M_{x1}, M_{x2} — средние арифметические аргумента (ряда x); $m_{y \cdot x1}^2, m_{y \cdot x2}^2$ — квадраты ошибок прогноза, или остаточные дисперсии.

Вначале вычислим усредненную сигму:

$$\sigma = \sqrt{\frac{(214 - 2) 49,47 + (7 - 2) 2,68}{214 + 7 - 4}} = 6,9564,$$

затем величину критерия Стьюдента (3.20):

$$t = \frac{(214 - 1) 35,868 \cdot 0,24826 + (7 - 1) 116,683 \cdot 0,308}{(214 - 1) 35,868 + (7 - 1) 116,683} - \frac{44,15 - 43,723}{75,047 - 75} =$$

$$= 6,9564 \sqrt{\frac{1}{(214 - 1) 35,868 + (7 - 1) 116,683} + \frac{1}{(75,047 - 75)^2} \left(\frac{1}{214} + \frac{1}{7} \right)} =$$

$$= -0,15536, \text{ или } |t| = 0,155.$$

Число степеней свободы для оценки величины t по табл. 3П определяется так же, как и в предыдущем пункте, по формуле: $\nu = N_1 + N_2 - 4$, т. е. $\nu = 214 + 7 - 4 = 217$. Табличная величина при этом числе степеней свободы и на доверительном уровне $P_1 = 95\%$ равна 1,960, что больше вычисленной величины критерия: $1,960 > 0,155$. Следовательно, различие между двумя сравниваемыми коэффициентами a_{01} и a_{02} уравнений случайное.

Таким образом, в результате проверки по трем этапам устанавливаем, что два сравниваемых уравнения прямой линии различаются несущественно и могут аналогично интерпретировать вид рассматриваемой зависимости. Для вывода о несущественном различии двух уравнений необходимо, чтобы выводы по всем трем этапам проверки совпали. Если хотя бы по одному пункту различие окажется существенным, то дальнейшая проверка может быть прекращена и уравнения должны считаться в целом достоверно различающимися.

§ 3.02. Уравнение множественной регрессии

Аналитическим выражением многофакторных связей являются уравнения множественной регрессии. В рассматриваемом здесь примере о высоте растений, весе 1000 семян и урожайности сои (табл. 3.04) связь между этими факторами оказалась в общем слабой. Более или менее целесообразным в данном случае было бы искать лишь параметры уравнения, выражающего зависимость веса семян от двух других факторов:

$$y = a_0 + a_1x + a_2z, \quad (3.21)$$

где y — функция, зависимая переменная, вес 1000 семян; a_0, a_1, a_2 — коэффициенты уравнения; x — высота растений (в см) и z — урожайность сои (в ц/га) — независимые переменные.

Первый способ. Для получения коэффициентов уравнения множественной регрессии по этому способу требуется решить систему

Таблица 3.04. Зависимость веса 1000 семян (y) от высоты растений (x) и урожайности (z) у сои (y' — теоретические значения веса семян)

	z , ц/га	y ,	y' ,	$x - M_x$	$z - M_z$	$y - M_y$
66	16	199	214,3	1,462	-4	-21,62
44	17	200	208,3	-20,538	-3	-20,62
77	15	210	216,5	12,462	-5	-10,62
63	20	212	220,1	-1,538	0	-8,62
62	19	217	218,0	-2,538	-1	-3,62
73	12	218	210,0	8,462	-8	-2,62
64	23	218	225,5	-0,538	3	-2,62
74	27	223	235,8	9,462	7	2,38
63	19	225	218,4	-1,538	-1	4,38
74	25	234	232,4	9,462	5	13,38
54	25	235	225,4	-10,538	5	14,38
43	25	235	221,5	-21,538	5	14,38
82	17	242	221,7	17,462	-3	21,38
				+0,006	0	0,06

трех нормальных уравнений, полученных по методу наименьших квадратов:

$$\Sigma y = a_0 N + a_1 \Sigma x + a_2 \Sigma z,$$

$$\Sigma yx = a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma zx, \quad (3.22)$$

$$\Sigma yz = a_0 \Sigma z + a_1 \Sigma xz + a_2 \Sigma z^2.$$

Необходимые для решения уравнений суммы вычислены в табл. 2.15: $\Sigma y = 2868$, $\Sigma x = 839$, $\Sigma yx = 185412$, $\Sigma x^2 = 55869$, $\Sigma yz = 57737$, $\Sigma xz = 16609$, $\Sigma z^2 = 5458$, $\Sigma z = 260$.

1. Подставляя эти суммы в систему уравнений (3.22), получим:

$$(1) \quad 2868 = 13a_0 + 839a_1 + 260a_2,$$

$$(2) \quad 185412 = 839a_0 + 55869a_1 + 16609a_2,$$

$$(3) \quad 57737 = 260a_0 + 16609a_1 + 5458a_2.$$

2. Решим систему, состоящую из первого и второго уравнений, для чего разделим первое уравнение на 13, а второе — на 839 и вычтем из второго первое:

$$\begin{array}{r} 220,9916 = a_0 + 66,589a_1 + 19,7962a_2 \\ - 220,6154 = a_0 + 64,5385a_1 + 20a_2 \\ \hline 0,3762 = 2,0515a_1 - 0,2038a_2 \end{array}$$

3. Чтобы получить второе уравнение с двумя неизвестными, решим систему, состоящую из первого и третьего уравнений, для чего разделим первое уравнение на 13, а третье — на 260 и вычтем

из третьего первое:

$$\begin{array}{r} 222,0654 = a_0 + 63,8808a_1 + 20,9923a_2 \\ - 220,6154 = a_0 + 64,5385a_1 + 20a_2 \\ \hline 1,4500 = \quad - 0,6577a_1 + 0,9923a_2 \end{array}$$

4. Теперь решим систему двух получившихся уравнений с двумя неизвестными:

$$\begin{array}{r} 0,3762 \quad 2,0515a_1 \quad - 0,2038a_2 \\ 1,4500 \quad -0,6577a_1 + 0,9923a_2, \end{array}$$

для чего разделим первое из них на 2,0515, а второе — на 0,6577 и, сложив полученные уравнения, найдем a_2 :

$$\begin{array}{r} + 0,183378 = a_1 - 0,099342a_2 \\ + 2,20465 = -a_1 + 1,50874a_2 \\ \hline 2,388028 = 1,409398a_2 \\ a_2 = 1,694367 \end{array}$$

5. Подставляя значение a_2 в первое уравнение из системы двух уравнений (пункт 4) с двумя неизвестными, определим a_1 :

$$\begin{aligned} 0,3762 &= 2,0515a_1 - 0,2038 \cdot 1,6944; \\ a_1 &= (0,3762 + 0,3453)/2,0515 = 0,3517. \end{aligned}$$

6. Подставляя в первое уравнение нашей начальной системы значения a_1 и a_2 , найдем a_0 :

$$\begin{aligned} 2868 &= 13a_0 + 839 \cdot 0,3517 + 260 \cdot 1,6944; \\ a_0 &= (2868 - 295,0763 - 440,5440)/13 = 164,030. \end{aligned}$$

7. Уравнение регрессии:

$$y_{xz} = 164,030 + 0,3517x + 1,69436z.$$

8. Проверка: подставляя в первое уравнение начальной системы значения a_0 , a_1 и a_2 , получаем:

$$\begin{aligned} 2868 &= 13 \cdot 164,030 + 839 \cdot 0,3517 + 260 \cdot 1,69436 = 2132,3900 + \\ &+ 295,0763 + 440,5336 = 2867,9999. \end{aligned}$$

Очевидно, что коэффициенты a_0 , a_1 , a_2 определены правильно.

9. Ошибка уравнения множественной регрессии определяется по формуле:

$$m_{y \cdot xz} = \sqrt{\frac{\sum a_y^2 - [a_1 \sum (a_y a_x) + a_2 \sum (a_y a_z)]}{N - n}} \quad (3.23)$$

где $m_{y \cdot xz}$ — ошибка теоретических значений функции y ; a_y , a_x , a_z — отклонения вариант рядов y , x , z от их средних арифметических; a_1 , a_2 — коэффициенты уравнения множественной регрессии; N — объем ряда; n — число коэффициентов уравнения, включая свободный член.

Суммы отклонений a_y, a_x, a_z вычисляются по формулам:

$$\begin{aligned}\Sigma a_y^2 &= \Sigma y^2 - NM_y^2, \\ \Sigma a_y a_x &= \Sigma yx - NM_y M_x, \\ \Sigma a_y a_z &= \Sigma yz - NM_y M_z.\end{aligned}\tag{3.24}$$

Пользуясь данными табл. 2.15, определяем средние арифметические рядов x, y, z по формуле (1.08):

$$M_x = 839/13 = 64,5384; \quad M_y = 2868/13 = 220,615;$$

$$M_z = 260/13 = 20,0;$$

$$\Sigma y^2 = 634906; \quad \Sigma yx = 185412; \quad \Sigma yz = 57737.$$

Подставляем найденные величины в формулы (3.24):

$$\Sigma a_y^2 = 634906 - 13 \cdot 220,615^2 = 2183,2831;$$

$$\Sigma a_y a_x = 185412 - 13 \cdot 220,615 \cdot 64,5384 = 316,1915;$$

$$\Sigma a_y a_z = 57737 - 13 \cdot 220,615 \cdot 20,0 = 377,100.$$

Теперь определены все величины, необходимые для вычисления ошибки уравнения множественной регрессии по формуле (3.23):

$$m_{y \cdot xz} = \sqrt{\frac{2183,2831 - (0,3517 \cdot 316,1915 + 1,6944 \cdot 377,1)}{13 - 3}} = \pm 11,971.$$

Таким образом, по уравнению $y_{xz} = 164,030 + 0,3517x + 1,6944z$, можно рассчитать вес 1000 семян, зная высоту растений и урожайность сои с ошибкой: $m_{y \cdot xz} = \pm 11,97$ г.

Второй способ. Систему нормальных уравнений (3.22) можно решить другим способом и получить коэффициенты множественного уравнения линейной регрессии следующим образом.

1. Находим по данным того же примера (табл. 2.15) те же суммы [они приведены вслед за уравнениями (3.22)] и средние арифметические всех трех рядов данных [они приведены к формулам (3.24)].

2. Определяем промежуточные величины:

$$A = M_x - \frac{\Sigma x^2}{\Sigma x}, \quad B = -\frac{\Sigma x^2}{\Sigma x} - \frac{\Sigma xz}{\Sigma z},\tag{3.25}$$

$$A = 64,53846 - \frac{55869}{839} = -2,05152,$$

$$B = \frac{55869}{839} - \frac{16609}{260} = 2,70921,$$

$$C = \frac{B}{A} = \frac{2,70921}{-2,05152} = -1,32059.$$

3. Находим коэффициент уравнения при z :

$$a_2 = \frac{C \left(M_y - \frac{\sum yx}{\sum x} \right) - \frac{\sum yx}{\sum x} + \frac{\sum yz}{\sum z}}{C \left(M_z - \frac{\sum zx}{\sum x} \right) - \frac{\sum zx}{\sum x} + \frac{\sum z^2}{\sum z}} \quad (3.26)$$

$$a_2 = \frac{-1,32059 \left(220,61538 - \frac{185412}{839} \right) - \frac{185412}{839} + \frac{57737}{260}}{-1,32059 \left(20 - \frac{16609}{839} \right) - \frac{16609}{839} + \frac{5458}{260}} =$$

$$= 1,69437.$$

4. Определяем коэффициент уравнения при x :

$$a_1 = \frac{M_x \sum y - \sum yx - a_2 (M_x \sum z - \sum zx)}{M_x \sum x - \sum x^2}, \quad (3.27)$$

$$a_1 = \frac{64,53846 \cdot 2868 - 185412 - 1,69437 (64,53846 \cdot 260 - 16609)}{64,53846 \cdot 839 - 55869} = 0,35174.$$

5. Находим свободный член уравнения множественной регрессии:

$$a_0 = M_y - a_1 M_x - a_2 M_z; \quad (3.28)$$

$$a_0 = 220,61538 - 0,35174 \cdot 64,53846 - 1,69437 \cdot 20 = 164,02709,$$

после чего все неизвестные коэффициенты определены и уравнение линейной зависимости веса 1000 семян (y) от высоты растения (x) и урожая сои (z) принимает конкретную форму:

$$y' = 164,027 + 0,35174x + 1,69437z.$$

Подставляя в это уравнение соответствующие величины x и z , получим теоретические значения веса 1000 семян (y'), приведенные в столбце 4 табл. 3.04; сумма их разностей с эмпирическими значениями (y), возведенных в квадрат $[\sum (y - y')^2]$, подставленная в формулу (3.23), дает величину ошибки прогноза по этому уравнению: $m_{y \cdot xz} = 11,9635$, т. е. примерно ту же, что и по первому способу. Данные в табл. 3.04 ранжированы по y , в отличие от табл. 2.15, где они не упорядочены.

Третий способ. По этому способу коэффициенты уравнения множественной регрессии находят через суммы квадратов и произведений отклонений от средней, что несколько сокращает расчеты, так как позволяет оперировать с меньшими по абсолютной величине числами. Данный, третий способ, может применяться в любых случаях, но особенно он рекомендуется тогда, когда исходные данные представлены многозначными числами, а также когда одновременно желательно вычислить дисперсии, сигмы всех трех рядов и коэффициенты корреляции между ними.

Рассмотрим технику расчетов по этому способу на том же числовом примере, что и в первых двух способах данного параграфа.

Таблица 3.05. Вычисление сумм для определения коэффициентов уравнения множественной регрессии по третьему способу

$(x - M_x)^2$	$(y - M_y)^2$	$(z - M_z)^2$	$(x - M_x)(y - M_y)$	$(z - M_z)(y - M_y)$	$(x - M_x)(z - M_z)$
1	2	3	4	5	6
2,1374	467,42	16	-31,608	86,48	-5,848
421,81	425,18	9	423,49	61,86	61,614
155,3	112,78	25	-132,35	53,1	-62,31
2,3654	74,304	0	13,258	0	0
6,4414	13,104	1	9,1876	3,62	2,538
71,605	6,8644	64	-22,17	20,96	-67,696
0,28944	6,8644	9	1,4096	-7,86	-1,614
89,529	5,6644	49	22,52	16,66	66,234
2,3654	19,184	1	-6,7364	-4,38	1,538
89,529	179,02	25	126,6	66,9	47,31
111,05	206,78	25	-151,54	71,9	-52,69
463,89	206,78	25	-309,72	71,9	-107,69
304,92	457,1	9	373,34	-64,14	-52,386
1721,2	2181,2	258	315,68	377	-171

1. Вычислим по данным табл. 3.04 средние арифметические наших трех рядов: $M_y = 220,62$; $M_x = 64,538$; $M_z = 20$; и следующие суммы (табл. 3.05): $\Sigma (x - M_x)^2 = 1721,2$; $\Sigma (y - M_y)^2 = 2181,2$; $\Sigma (z - M_z)^2 = 258$; $\Sigma (x - M_x)(y - M_y) = 315,68$; $\Sigma (z - M_z)(y - M_y) = 377$; $\Sigma (x - M_x)(z - M_z) = -171$.

2. Найдем вспомогательные величины:

$$B = \frac{\Sigma (z - M_z)^2}{\Sigma (x - M_x)(z - M_z)}, \quad (3.29)$$

$$B = 258 / -177 = -1,5088;$$

$$A = B \Sigma (x - M_x)^2 - \Sigma (x - M_x)(z - M_z); \quad (3.30)$$

$$A = -1,5088 \cdot 1721,2 + 171 = -2426.$$

3. Коэффициент при x равен:

$$a_1 = [B \Sigma (x - M_x)(y - M_y) - \Sigma (z - M_z)(y - M_y)] / A, \quad (3.31)$$

$$a_1 = [-1,5088 \cdot 315,68 - 377] / -2426 = 0,35173.$$

4. Коэффициент при z равен:

$$a_2 = \frac{1}{A} \left[\frac{\Sigma (x - M_x)^2 \cdot \Sigma (z - M_z)(y - M_y)}{\Sigma (x - M_x)(z - M_z)} - \Sigma (x - M_x)(y - M_y) \right], \quad (3.32)$$

$$a_2 = \frac{1}{-2426} \left[\frac{1721,2 \cdot 377}{-171} - 315,68 \right] = 1,6944.$$

5. Свободный член уравнения по формуле (3.28) равен:

$$a_0 = 220,62 - 0,3517 \cdot 64,538 - 1,6944 \cdot 20 = 164,03,$$

т. е. величины всех коэффициентов совпадают с полученными выше. Используя суммы из табл. 3.05, можно также вычислить сигмы (1.24) всех трех рядов и коэффициенты корреляции между ними (2.18):

$$\sigma_x^2 = \frac{1721,2}{13-1} = 143,43; \quad \sigma_x = 11,976;$$

$$\sigma_y^2 = \frac{2181,2}{13-1} = 181,77; \quad \sigma_y = 13,482;$$

$$\sigma_z^2 = \frac{258}{13-1} = 21,5; \quad \sigma_z = 4,6369;$$

$$r_{xy} = \frac{315,68}{\sqrt{1721,2 \cdot 2181,2}} = 0,16292;$$

$$r_{xz} = \frac{-171}{\sqrt{1721,2 \cdot 258}} = -0,2566;$$

$$r_{zy} = \frac{377}{\sqrt{258 \cdot 2181,2}} = 0,50254.$$

Величины коэффициентов корреляции между этими же признаками совпадают с вычисленными в § 2.16 по другому способу.

Оценку достоверности значений найденных коэффициентов, или, точнее, оценку достоверности их отличия от нуля, делаем при помощи формул:

$$t(a_1) = \frac{a_1}{m_{y \cdot xz}} \sqrt{\sum (x - M_x)^2 (1 - r_{xz}^2)}, \quad (3.33)$$

$$t(a_2) = \frac{a_2}{m_{y \cdot xz}} \sqrt{\sum (z - M_z)^2 (1 - r_{xz}^2)}, \quad (3.34)$$

где $t(a_1)$, $t(a_2)$ — величины критерия Стьюдента, сравниваемые с табличными величинами (табл. 3П), при числе степеней свободы: $v = N - 3$; $m_{y \cdot xz}$ — ошибка уравнения по формуле (3.23); $\sum (x - M_x)^2$, $\sum (z - M_z)^2$ — суммы квадратов отклонений величин x , z из табл. 3.05; r_{xz} — коэффициент корреляции между рядами x , z .

Величина критерия Стьюдента соответственно для коэффициентов a_1 и a_2 равна:

$$t(a_1) = \frac{0,35174}{11,963} \sqrt{1721,2 (1 - 0,2566^2)} = 1,179;$$

$$t(a_2) = \frac{1,6944}{11,963} \sqrt{258 (1 - 0,2566^2)} = 2,199.$$

По табл. 3П при числе степеней свободы $v = 13 - 3 = 10$ и на 95 % доверительном уровне $t = 2,228$, что больше обеих вычисленных величин $t(a_1)$ и $t(a_2)$; следовательно, вес 1000 семян сои (y) существ-

венно не зависит ни от средней высоты растений (x), ни от урожайности (z). Причиной того, что влияние этих факторов на вес семян в данном случае не доказано, является недостаточное число наблюдений.

В заключение заметим, что при исследовании комплексов взаимосвязанных признаков можно установить структуру их связей различными способами. Наиболее простым из них является способ троек признаков, когда исследуются взаимосвязи комплекса признаков во всех возможных их сочетаниях по три, а затем устанавливается структура связей по максимальным величинам корреляционных отношений и коэффициентов корреляции (Зайцев, 1978). При этом исключаются многие побочные эффекты взаимодействия, о которых в настоящее время пока еще мало что известно, и сфера исследований не выходит за пределы реально доступного нам трехмерного пространства.

§ 3.03. Парабола второго порядка

Подбор соответствующей теоретической криволинейной функции к данной эмпирической зависимости пока не формализован и осуществляется главным образом на основании соответствующего опыта исследователя методом проб и ошибок. Наибольшую помощь при подборе кривых оказывает сборник их графиков (Семендяев, 1933), воспроизведенных в § 3.14.

В биологии нередко применяется аппроксимация криволинейных эмпирических зависимостей параболой второго порядка, уравнение которой в общей форме:

$$y' = a_0 + a_1x + a_2x^2, \quad (3.35)$$

где y' — функция, или зависимая переменная; x — аргумент, или независимая переменная; a_0, a_1, a_2 — коэффициенты уравнения. Вычисление конкретных значений коэффициентов уравнения (3.35) и является содержанием работы по аппроксимации данной эмпирической зависимости параболой второго порядка.

Попробуем посредством параболы отобразить зависимость между высотой растения и урожаем семян с одного растения у кориандра. Первичные данные были обработаны с построением корреляционной решетки (которая здесь не приводится) и вычислением эмпирической линии регрессии (по способу, приведенному в § 2.01), ее значения и приняты в качестве исходных данных (табл. 3.06) для определения коэффициентов параболы, которые вычисляются разными способами.

Первый способ. При использовании метода наименьших квадратов в его обычной форме, для нахождения коэффициентов уравнения (3.35) требуется решить систему нормальных уравнений:

$$\begin{aligned} a_0N + a_1\sum x + a_2\sum x^2 &= \sum y, \\ a_0\sum x + a_1\sum x^2 + a_2\sum x^3 &= \sum yx, \\ a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 &= \sum yx^2, \end{aligned} \quad (3.36)$$

Таблица 3.06. Вычисление сумм для определения коэффициентов уравнения параболы второго порядка методом наименьших квадратов

Высота растений корнандра, см, x				Вес семян с одного растения, г, y	xy	yx^2	y'
112	12 544	1 404 928	157 351 936	24,0	2688,0	301 056,0	22,97
95	9 025	857 375	81 450 625	17,6	1672,0	158 840,0	18,69
78	6 084	474 552	37 015 056	13,2	1029,6	80 308,8	15,14
61	3 721	226 981	13 845 841	14,4	878,4	53 582,4	12,34
44	1 936	85 184	3 748 096	11,2	492,8	21 683,2	10,29
27	729	19 683	531 441	8,0	216,0	5 832,0	8,97
417	34 039	3 068 703	293 942 995	88,4	6976,8	621 302,4	88,40

где a_0, a_1, a_2 — искомые коэффициенты уравнения (3.35); N — число точек эмпирической линии регрессии; $\Sigma x, \Sigma x^2, \Sigma x^3, \Sigma x^4, \Sigma y, \Sigma xy, \Sigma yx^2$ — суммы, полученные в результате вычислений, выполненных в табл. 3.06.

Подставляя найденные суммы в уравнения (3.36), получаем:

$$6a_0 + 417a_1 + 34039a_2 = 88,4;$$

$$416a_0 + 34039a_1 + 3068703a_2 = 6976,8;$$

$$34039a_0 + 3068703a_1 + 293942995a_2 = 621302,4.$$

Решая эту систему уравнений относительно a_0, a_1 и a_2 , находим: $a_0 = 8,41$; $a_1 = -0,013937$; $a_2 = 0,0012852$. Уравнение (3.35) в конкретном выражении имеет вид: $y' = 8,41 - 0,013937x + 0,0012852x^2$, где y' — теоретические значения функции, т. е. веса семян с одного растения (в г); x — аргумент, т. е. высота растений у корнандра (в см).

В табл. 3.06 в последней колонке вычислены теоретические значения функции. Подобно тому, как это сделано в табл. 3.01, вычислим попарно разности $(y - y')$, возведем все разности в квадрат и получим их сумму:

$$\Sigma (y - y')^2 = 12,025.$$

Применив к уравнению параболы формулу (3.04), определим его ошибку:

$$m_{y \cdot x} = \sqrt{\frac{12,025}{6 - 3}} = 2,0 \text{ г.}$$

Точки пересечения параболы с осью ординат и осью абсцисс соответственно равны:

$$y = a_0, \quad x = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_2}. \quad (3.37)$$

По формуле (3.37) подкоренное выражение (дискриминант) для данных по рассматриваемому примеру равно $-0,043$, поэтому точек пересечения с осью абсцисс у этой параболы не существует.

Абсцисса точки максимума (или минимума) параболы равна

$$x = -\frac{a_1}{2a_2}, \quad (3.38)$$

или для нашего примера:

$$x = -\frac{-0,013937}{2 \cdot 0,0012852} = 5,42.$$

Обозначения коэффициентов в формулах (3.37) и (3.38) те же, что и в формуле (3.35). Как видно из рис. 18, совпадение эмпирической линии регрессии с теоретической по параболе второго порядка не совсем полное. В § 3.04 показано, каким образом в подобных случаях можно улучшить качество аппроксимации, повышая порядок параболы.

Второй способ. Те же вычисления, которые сделаны по первому способу, можно представить в несколько более компактной форме, заменив приведенное решение системы нормальных уравнений формулами:

$$a_2 = \frac{N(te - hc) + b(hb - de) + c(dc - fb)}{c(c^2 - 2be - Ng) + Ne^2 + gb^2}, \quad (3.39)$$

$$a_1 = \frac{db - fN - a_2(cb - Ne)}{b^2 - Nc}, \quad (3.40)$$

$$a_0 = \frac{d - a_1b - a_2c}{N}, \quad (3.41)$$

где $b = \sum x = 417$; $c = \sum x^2 = 34039$; $d = \sum y = 88,4$; $e = \sum x^3 = 3068703$; $f = \sum yx = 6976,8$; $g = \sum x^4 = 293942995$; $h = \sum yx^2 = 621302,4$; N — объем совокупности; a_0, a_1, a_2 — коэффициенты уравнения.

Подставив указанные числовые значения из табл. 3.06 в формулы (3.39)–(3.41), получим величины: $a_2 = 0,00128$; $a_1 = -0,01395$; $a_0 = 8,41124$, которые практически совпадают с найденными по первому способу. Величины ошибок уравнения, дискриминанта уравнения, точки максимума уравне-

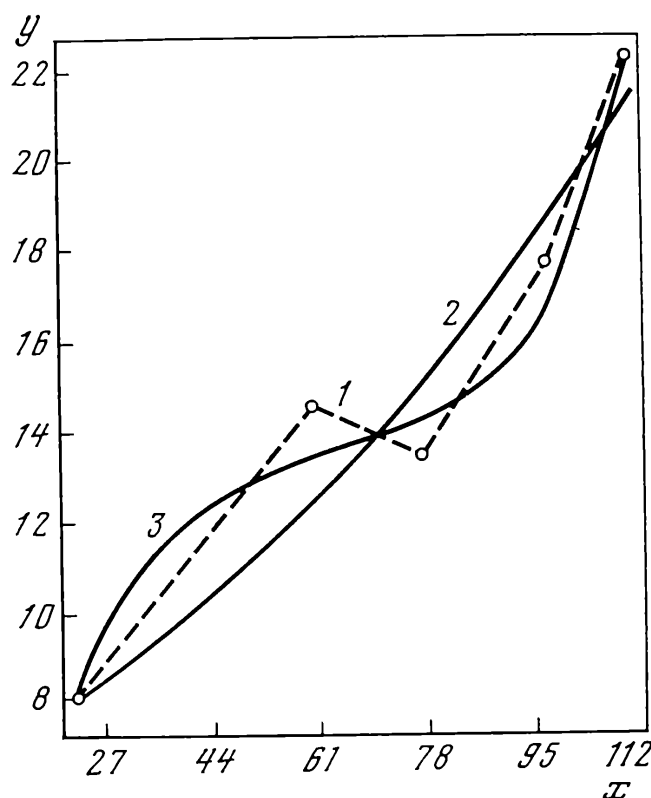


Рис. 18. Связь между высотой растения кориандра (x) и урожаем семян с него (y)

Эмпирическая (1) и теоретические линии регрессии: 2 — парабола второго порядка, 3 — парабола третьего порядка

ния по формулам (3.04), (3.37), (3.38) также равны найденным выше. Способы первый и второй целесообразно применять в случае мало-значных величин x и y .

Третий способ. Этот способ целесообразно применять, когда величины x многозначные, а величины y — мало-значные, шаг интервала в ряду x может быть как одинаковым, так и неравным во всех интервалах. Алгоритм заключается в следующем (числовой пример по тем же данным из табл. 3.06).

1. Вычислим среднюю арифметическую ряда x :

$$M_x = (\Sigma x)/N = 417/6 = 69,5,$$

2. Кодировем варианты ряда x по формуле: $x_0 = x - M_x$, получим: 42,5; 25,5; 8,5; —8,5; —25,5; —42,5.

3. Вычислим суммы: $d = \Sigma y = 88,4$; $h = \Sigma yx_0^2 = 78521,3$; $c = \Sigma x_0^2 = 5057,5$; $g = \Sigma x_0^4 = 7381168,375$; $f = \Sigma yx_0 = 833$; $N = 6$.

Для вычисления этих величин при ручном счете следует составить таблицу, построение которой ясно из перечня сумм, и повторить расчеты два раза.

4. Искомые коэффициенты в соответствии с обозначениями пунктов 1—3 найдем по формулам:

$$a_2 = \frac{d - \frac{Nh}{c}}{c - \frac{Ng}{c}}, \quad a_1' = \frac{f}{c}, \quad a_0' = \frac{d - a_2c}{N}, \quad (3.42)$$

$$a_1 = a_1' - 2a_2M_x, \quad a_0 = a_0' + M_x(M_xa_2 - a_1'), \quad (3.43)$$

их величины совпадают с найденными выше; $a_1' = 0,1647$; $a_0' = 13,65$. Квадраты разностей эмпирических и теоретических значений функции в данном примере $(y - y')^2$ равны: 1,058; 2,237; 6,011; 10,243; 11,079, сумма этих величин равна 12,023 и ошибка по формуле (3.04) равна 2 г, т. е. той же величине, что и по первому способу.

Четвертый способ. Этот способ как наиболее рациональный рекомендуется применять в большинстве расчетов коэффициентов парабол, но особенно в тех случаях, когда значения как x , так и y представлены двух-, трех и многозначными числами. Желательно, чтобы варианты ряда x были отделены одинаковым шагом интервала, как в нашем примере, где они следуют через $c = 17$. В табл. 3.07 вычислены суммы, необходимые для дальнейших расчетов.

1. В столбце 2 варианты ряда x кодированы по формуле, приведенной в его заголовке, где A — начальная (минимальная) варианта ряда; c — классовый интервал ряда x , равный в данном случае 17.

2. Средняя арифметическая величин из столбца 2 равна $M_x = (\Sigma x')/N = 15/6 = 2,5$, где N — объем ряда.

3. Из величин столбца 2 вычитаем их среднюю: $x_0 = x' - M_x$, и результаты записываем в столбец 3.

4. Средняя арифметическая ряда y равна: $M_y = (\Sigma y)/N = 88,4/6 = 14,73$.

Таблица 3.07. К вычислению коэффициентов параболы по четвертому способу ($A = 27$, $c = 17$)

	$x' = \frac{x - A}{c}$	$x_0 = x' - M_x$	y	$y_0 = y - M_y$	x_0^2	$y_0 x_0^2$	$y_0 x_0$	x_0^4
1	2	3	4	5	6	7	8	9
112	5	2,5	24,0	9,27	6,25	57,94	23,18	39,0625
95	4	1,5	17,6	2,87	2,25	6,46	4,31	5,0625
78	3	0,5	13,2	-1,53	0,25	-0,38	-0,76	0,0625
61	2	-0,5	14,4	-0,33	0,25	-0,08	0,17	0,0625
44	1	-1,5	11,2	-3,53	2,25	-7,94	5,29	5,0625
27	0	-2,5	8,0	-6,73	6,25	-42,06	16,83	39,0625
	15	0	88,4	+0,02	17,5	+13,94	+49,02	88,375

5. Из величин столбца 4 вычитаем их среднюю: $y_0 = y - M_y$ и результат записываем в столбце 5.

6. В столбце 6 записаны квадраты чисел из столбца 3.

7. В столбце 7 — произведения чисел из столбцов 5 и 6.

8. В столбце 8 — произведения чисел столбцов 3 и 5.

9. В столбце 9 — квадраты чисел столбца 6.

10. Суммируем числа столбцов 2—9.

11. Вычисляем первые промежуточные величины трех коэффициентов параболы:

$$a'_2 = \frac{\sum y_0 x_0^2}{\sum x_0^4 - (\sum x_0^2)^2 N^{-1}} = \frac{13,94}{88,375 - \frac{1}{6} (17,5)^2} = 0,3734; \quad (3.44)$$

$$a'_1 = \frac{\sum y_0 x_0}{\sum x_0^2} = \frac{49,02}{17,5} = 2,8011; \quad (3.45)$$

$$a'_0 = -\frac{a'_2 \sum x_0^2}{N} = -\frac{0,3734 \cdot 17,5}{6} = -1,0891. \quad (3.46)$$

12. Вычисляем вторые промежуточные величины двух коэффициентов:

$$a''_1 = a'_1 - 2a'_2 M_x = 2,8011 - 2 \cdot 0,3734 \cdot 2,5 = 0,9341, \quad (3.47)$$

$$a''_0 = M_y + a'_0 + M_x (a'_2 M_x - a'_1) = 14,73 - 1,0891 + 2,5 \cdot (0,3734 \cdot 2,5 - 2,8011) = +8,9719. \quad (3.48)$$

Вторую промежуточную величину коэффициента a_2 вычислять не надо, так как она совпадает с его первой промежуточной величиной. Полученные величины коэффициентов на этом этапе вычислений уже можно использовать для нахождения теоретических значений функции при условии, что в уравнение будут подставлены кодированные величины аргумента: $x' = (x - A)/c$, в нашем примере приведенные в столбце 2 табл. 3.07. Подставляя в уравнение: $y' = 8,9719 +$

+ $0,9341x' + 0,3734 (x')^2$ значения x' , получим значения y' , практически совпадающие с приведенными в последнем столбце табл. 3.06.

13. Если требуется получить уравнение, пригодное для вычислений y' по исходным значениям x , то производится дальнейшее преобразование коэффициентов:

$$a_2 = \frac{a_2'}{c^2} = \frac{0,3734}{17^2} = 0,001292; \quad (3.49)$$

$$a_1 = \frac{a_1''}{c} - 2a_2A = \frac{0,9341}{17} - 2 \cdot 0,001292 \cdot 27 = -0,01482; \quad (3.50)$$

$$\begin{aligned} a_0 &= a_0'' - A \left(\frac{a_1''}{c} - a_2A \right) = \\ &= 8,9719 - 27 \left(\frac{0,9341}{17} - 0,001292 \cdot 27 \right) = 8,4302, \end{aligned} \quad (3.51)$$

что дает уравнение следующего вида:

$$y' = 8,4302 - 0,01482x + 0,001292x^2,$$

т. е. то же, которое было получено по способам первому — третьему.

В § 3.06 приводятся также алгоритмы вычисления уравнений парабол первого, второго и третьего порядков по способу сумм.

Определение точек пересечения парабол второго порядка. При определении зоны оптимума для наибольшего накопления массы растительности в связи с некоторыми вопросами интродукции предполагается, что эта зона расположена около термического экватора (ТЭ) земли, где аборигенные растения находятся в зоне комфорта, так как обеспечены максимальным количеством света и тепла. Однако положение ТЭ известно до сих пор весьма приблизительно, и количественные данные об этой важной геофизической константе земли расходятся. Здесь приводится методика расчета точки пересечения двух парабол, пригодная для определения критических точек кривых двух процессов, переходящих один в другой, в том числе широтного изменения средней температуры в северном и южном полушариях земли. Исходя из уравнений парабол: $y_1 = a_1 + b_1x + c_1x^2$ и $y_2 = a_2 + b_2x + c_2x^2$, была найдена формула вычисления абсциссы точки их пересечения:

$$x_0 = \frac{(b_2 - b_1) \pm \sqrt{(b_1 - b_2)^2 - 4(c_1 - c_2)(a_1 - a_2)}}{2(c_1 - c_2)}. \quad (3.52)$$

Ординату (y_0) точки пересечения этих парабол получим, подставив x_0 в любое из двух исходных уравнений.

Приведем пример расчета. По данным справочной литературы среднегодовая температура воздуха (в °C) на соответствующих градусах широты равна: в южном полушарии: 90° ю. ш. — минус 32,6, 80° — минус 21,9, 70° — минус 12,2, 60° — минус 3,4, 50° — 4,4, 40° — 11,3, 30° — 17,2, 20° — 22,1, 10° — 26,1, в северном полушарии: 10° с. ш. — 28,2, 20° — 23,6, 30° — 18,5, 40° — 12,7, 50° — 6,4, 60° — минус 0,54, 70° — минус 8, 80° — минус

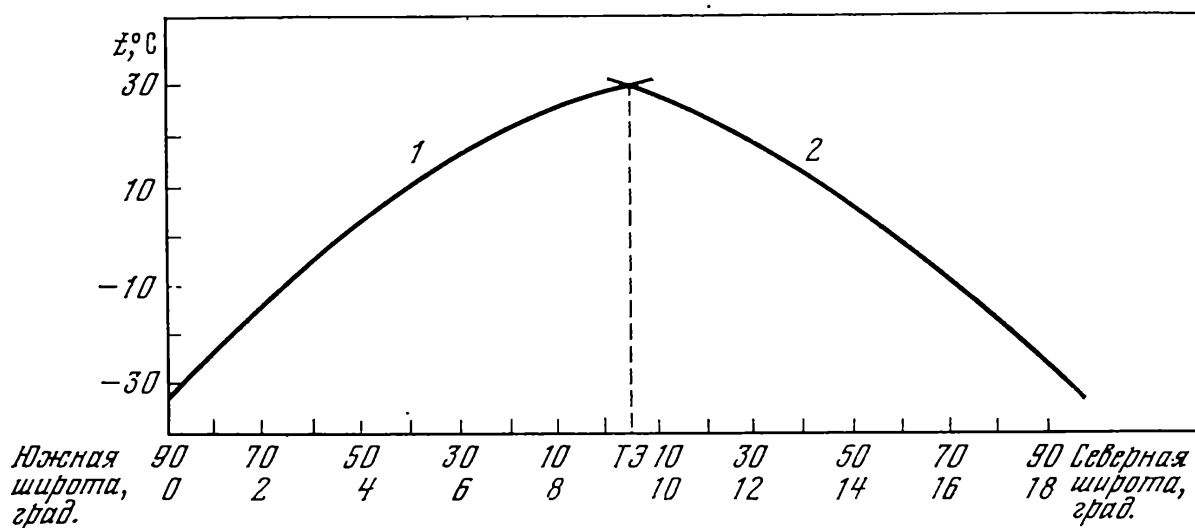


Рис. 19. Определение точки пересечения парабол 1 и 2

ТЭ — термический экватор

16,1, 90° — минус 24,8. Коэффициенты парабол по этим данным равны для южного полушария: $a_1 = -32,606$; $b_1 = 11,173$; $c_1 = -0,47929$; для северного полушария: $a_2 = 41,619$; $b_2 = 1,5905$; $c_2 = -0,29332$. Подставляя величины найденных коэффициентов в формулу (3.52), получим абсциссу: $x_0 = 4,97^\circ \approx 5^\circ$ с. ш., что соответствует положению ТЭ по использованным данным. Причина, вызывающая более холодный климат южного полушария известна — это преобладание в нем океанических масс; в северном полушарии выше, чем в южном, процент суши от общей его площади. Найденная величина ТЭ не является постоянной, она колеблется в зависимости от различных циклов изменения средней температуры воздуха на земле. В свою очередь систематическое, например ежегодное, определение положения ТЭ могло бы помочь в идентификации тех или иных сдвигов параметров климата земли. Преимуществом предлагаемого метода определения ТЭ являются его точность и учет одновременного влияния метеонаблюдений из всех регионов земли, использованных для расчетов ТЭ.

На рис. 19 пунктиром показан перпендикуляр, опущенный из точки пересечения парабол на ось абсцисс, где точка ТЭ соответствует положению термоэкватора. Уравнения линий 1 и 2 дают значения среднегодовой температуры воздуха при подстановке в них x , равных для 1 уравнения от 0 до 9, а для 2 уравнения — от 9 до 18. Эти условные градусы широты, приведенные также на рисунке, соответствуют значениям через 10° от 90° ю. ш. до 90° с. ш. Экватору соответствует при этом условная широта 9.

§ 3.04. Парабола третьего порядка

Если при аппроксимации криволинейной зависимости оказалось, что парабола второго порядка неудовлетворительно совпадает с эмпирической линией регрессии, то можно попытаться улучшить это совпадение, применяя уравнение параболы третьего порядка (или

кубической параболы):

$$y' = a_0 + a_1x + a_2x^2 + a_3x^3, \quad (3.53)$$

где y' — функция, или зависимая переменная; $a_0 — a_3$ — коэффициенты уравнения; x — аргумент, или независимая переменная.

Для определения коэффициентов уравнения (3.53) можно применить метод наименьших квадратов, однако вычисления при этом становятся слишком громоздкими. Поэтому для нахождения коэффициентов уравнения кубической параболы рекомендуется применять метод Чебышева, или способ сумм (§ 3.06).

Рассмотрим вычисление коэффициентов уравнения по методу Чебышева на примере из предыдущего параграфа. Особенность этой схемы расчета состоит в том, что варианты ряда аргумента должны быть расположены через равные интервалы, т. е. с одинаковым шагом.

1. Эмпирическая линия регрессии в данном случае состоит из 6 точек. В табл. 14П приведены числа Чебышева, при помощи которых можно вычислить коэффициенты уравнений параболы первого (прямая линия), второго и третьего порядка. Из столбца 6 (число точек эмпирической линии регрессии) табл. 14П выписываем числа Чебышева (j_1, j_2, j_3) и их суммы в табл. 3.08.

2. Умножаем значения y поочередно на соответствующие им j_1, j_2 и j_3 ; получаем в шестой строке сверху: $8,0 \cdot (-5) = -40,0$; $11,2 \cdot (-3) = -33,6$ и т. д. Произведения yj_1, yj_2, yj_3 суммируем по строкам.

3. Находим условные значения коэффициентов уравнения:

$$\begin{aligned} a'_0 &= \frac{\sum y}{N} = \frac{88,4}{6} = 14,733; \\ a'_1 &= \frac{\sum yj_1}{\sum j_1^2} = \frac{98,0}{70} = 1,40; \\ a'_2 &= \frac{\sum yj_2}{\sum j_2^2} = \frac{20,8}{84} = 0,2476; \\ a'_3 &= \frac{\sum yj_3}{\sum j_3^2} = \frac{40,0}{180} = 0,2222. \end{aligned} \quad (3.54)$$

4. Перемножаем условные коэффициенты a'_1, a'_2, a'_3 на соответствующие числа j_1, j_2, j_3 и результаты записываем в строках 9, 10, 11 (сверху): $1,4 \cdot (-5) = -7,0$; $1,4 \cdot (-3) = -4,2$ и т. д.

5. Вычисляем теоретические величины функции, для чего суммируем по вертикали (в строках 9, 10, 11) следующие числа:

$$y' = a'_0 + a'_1j_1 + a'_2j_2 + a'_3j_3,$$

например: $y' = 14,733 - 7,0 + 1,238 - 1,111 = 7,86$;

$$y' = 14,733 - 4,2 - 0,2476 + 1,5554 = 11,8408 \text{ и т. д.}$$

Таблица 3.08. Вычисление коэффициентов параболы третьего порядка по методу Чебышева

Высота растений, см, x	27	44	61	78	95	112	Суммы
Вес семян с 1 ра- стения, г, y	8,0	11,2	14,4	13,2	17,6	24,0	$\sum y = 88,4$
i_1	-5	-3	-1	+1	+3	+5	$\sum i_1^2 = 70$
i_2	+5	-1	-4	-4	-1	+5	$\sum i_2^2 = 84$
i_3	-5	+7	+4	-4	-7	+5	$\sum i_3^2 = 180$
yi_1	-40,0	-33,6	-14,4	+13,2	52,8	120,0	$\sum yi_1 = 98,0$
yi_2	40,0	-11,2	-57,6	-52,8	-17,6	120,0	$\sum yi_2 = 20,8$
yi_3	-40,0	78,4	57,6	-52,8	-123,2	120,0	$\sum yi_3 = 40,0$
$a'_1 i_1$	-7,0	-4,2	-1,4	1,4	4,2	7,0	
$a'_2 i_2$	1,238	-0,2476	-0,9904	-0,9904	-0,2476	1,238	
$a'_3 i_3$	-1,111	1,5554	0,8888	-0,8888	-1,5554	1,111	
y'	7,83	11,87	13,32	14,43	17,44	24,58	
$ y - y' $	0,17	0,67	1,08	1,23	0,16	0,50	
$(y - y')^2$	0,0289	0,4489	1,1664	1,5129	0,0256	0,2500	$\sum = 3,4327$

Таблица 3.09. Таблица приращений для вычисления коэффициентов уравнения кубической параболы по методу Чебышева

$\Delta'y'$	$\Delta'x$	$\Delta'x^2$	$\Delta'x^3$	$\Delta''y'$	$\Delta''x^2$	Δ''	$\Delta'''y'$	Δ'''
1	2	3	4	5	6	7	8	9
1,39	17	1785	141 797	—0,37	578	105 774	2,23	29 478
1,02	17	2363	247 571	1,86	578	135 252	—	—
2,88	17	2941	382 823	—	—	—	—	—

6. Для определения величин истинных коэффициентов уравнения выпишем 4 любые равноотстоящие (или рядом стоящие) величины y' и x и образуем систему уравнений следующего вида:

$$11,84 = a_0 + 44a_1 + 1936a_2 + 85184a_3,$$

$$13,23 = a_0 + 61a_1 + 3721a_2 + 226981a_3,$$

$$14,25 = a_0 + 78a_1 + 6084a_2 + 474552a_3,$$

$$17,13 = a_0 + 95a_1 + 9025a_2 + 857375a_3.$$

Числа при a_2 и a_3 представляют собой соответственно квадраты и кубы рядом стоящих величин x : 44, 61, 78, 95.

Вычитая предыдущее значение y' из последующего, образуем столбец 1 табл. 3.09: 13,23 — 11,84 = 1,39 и т. д. В столбце 2 записаны разности величин аргумента при коэффициенте a_1 : 61 — 44 = 17; 78 — 61 = 17; 95 — 78 = 17. В столбце 3 — разности чисел: последующее минус предыдущее при коэффициенте a_2 ; то же самое записано в столбце 4 для чисел при a_3 (см. систему четырех уравнений выше). В столбце 5 разности чисел из столбца 1, в 6 — разности чисел столбца 3, в 7 — разности чисел столбца 4, в 8 — разности чисел столбца 5, в 9 — разности чисел столбца 7. Во всех случаях в табл. 3.09 вычитается вышестоящее число из нижестоящего.

7. Истинные коэффициенты уравнения находим, решая следующие выражения:

$$a_3 = \frac{\Delta'''y'}{\Delta'''x^3} = \frac{2,23}{29478} = 0,00007565,$$

$$a_2 = \frac{\Delta''y' - a_3\Delta''x^3}{\Delta''x^2} = \frac{-0,37 - 0,00007565 \cdot 105774}{578} = -0,014512,$$

$$\begin{aligned} a_1 &= \frac{\Delta'y' - a_2\Delta'x^2 - a_3\Delta'x^3}{\Delta'x} = \\ &= \frac{1,39 + 0,014512 \cdot 1785 - 0,00007565 \cdot 141797}{17} = 0,974528, \end{aligned} \quad (3.55)$$

$$\begin{aligned} a_0 &= y' - a_1x - a_2x^2 - a_3x^3 = 11,84 - 0,974528 \cdot 44 + \\ &+ 0,014512 \cdot 1936 - 0,00007565 \cdot 85184 = -9,3882. \end{aligned}$$

Уравнение (3.53) в результате получает следующую конкретную форму:

$$y' = -9,39 + 0,974x - 0,0145x^2 + 0,000076 x^3.$$

Для вычисления теоретических значений функции уравнение (3.53) лучше представить в форме:

$$y' = a_0 + x [a_1 + x (a_2 + a_3x)].$$

8. Теоретические значения функции, их разности с фактическими значениями ($y - y'$) и квадраты последних (табл. 3.08) необходимы для вычисления ошибки прогноза по данному уравнению по формуле (3.04):

$$m_{y \cdot x} = \sqrt{\frac{3,4327}{6-4}} = 1,7.$$

Линия 3 на рис. 18 соответствует уравнению параболы третьего порядка и ближе отражает эмпирическую зависимость (линия 1), чем парабола второго порядка (линия 2). Ошибка теоретического уравнения регрессии также уменьшилась: с 2,0 (см. § 3.03) до 1,7.

Методика вычисления по способу Чебышева коэффициентов уравнений парабол, которые заданы рядами с неодинаковым шагом по аргументу, приводится в книге В. И. Хотимского (1959).

Точка пересечения параболы третьего порядка с осью ординат равна $y = a_0$, абсцисса ее точки максимума:

$$x = \frac{-a_2 - \sqrt{a_2^2 - 3a_3a_1}}{3a_3}, \quad (3.56)$$

точка перегиба параболы:

$$x = -\frac{a_2}{3a_3}, \quad (3.57)$$

где обозначения те же, что и к формуле (3.53).

В рассматриваемом примере парабола не пересекается с осью абсцисс, так как при подстановке значений коэффициентов в формулу (3.56) в подкоренном выражении получается отрицательное число. Абсцисса точки перегиба:

$$x = -\frac{-0,0145}{3 \cdot 0,000076} = 63,6 \text{ см.}$$

§ 3.05. Обращенная парабола

Зависимость изменений интенсивности физиологических функций или морфологических особенностей организма от количественного изменения какого-либо экологического фактора, например тепла, нередко можно выразить кривой, уравнение которой

$$y' = \frac{x}{a_0 + a_1x + a_2x^2}, \quad (3.58)$$

где y' — теоретическое значение функции; x — аргумент; a_0 , a_1 , a_2 — коэффициенты уравнения.

Свое название эта кривая получила в связи с тем, что она легко может быть приведена к уравнению параболы второго, третьего или более высокого порядка делением на x обеих частей уравнения с одновременным их обращением, после чего уравнение (3.58) принимает вид: $x/y = a_0 + a_1x + a_2x^2$; обозначив $y_0 = x/y$, получим уравнение параболы второго порядка.

Конкретные значения коэффициентов можно находить способом наименьших квадратов (см. второй способ), однако для той же цели можно применить также более простой метод средних.

Первый способ. Посредством метода средних выровнена зависимость прироста (в мм) грибницы плесневого гриба *Coniophora cerebella* от температуры. Измерения длины грибницы производились через каждые двое суток. Вычисления, приведенные в табл. 3.10, выполняются в следующем порядке.

1. Величины x делим на соответствующие им y (столбец 3).

2. Выбираем пару значений x_1 (из столбца 1), $(x/y)_1$ (из столбца 3) в качестве избранной точки [в данном случае $x_1 = 15$; $(x/y)_1 = 0,9375$] и поочередно вычитаем число 15 из всех чисел столбца 1 и число 0,9375 из всех чисел столбца 3.

3. Числа столбца 5 поделим на соответствующие им числа столбца 4.

4. Для нахождения коэффициентов a_1 и a_2 применяются уравнения:

$$\begin{aligned}\sum_1 \frac{y_a}{x_a} &= n_1 a'_1 + a_2 \sum_1 x, \\ \sum_2 \frac{y_a}{x_a} &= n_2 a'_1 + a_2 \sum_2 x, \\ a'_1 &= a_1 + x_1 a_2,\end{aligned}\tag{3.59}$$

где $\sum_1 \frac{y_a}{x_a}$ — сумма вариант верхней части ряда в столбце 6, в данном случае она равна 0,23484; $\sum_2 \frac{y_a}{x_a}$ — сумма нижней, оставшейся части того же ряда из столбца 6, в данном случае она равна 1,05854; $n_1 = 5$ — число вариант, суммированных в верхней части столбца 6; $n_2 = 4$ — число вариант, суммированных в нижней части столбца 6; $\sum_1 x = 75$ — сумма соответствующих вариант верхней части ряда x из столбца 1; $\sum_2 x = 150$ — сумма соответствующих вариант нижней части ряда из столбца 1; a'_1 — условное значение коэффициента a_1 ; a_1, a_2 — коэффициенты уравнения (3.58). Заметим, что ряды в столбцах 1 и 6 могли бы быть разделены на любые две части; лучше, если они будут приблизительно равными; $x_1 = 15$ — абсцисса избранной точки.

Подставляя найденные суммы в уравнения (3.59), получаем:

$$0,23484 = 5a'_1 + 75a_2; \quad 1,05854 = 4a'_1 + 150a_2.$$

Таблица 3.10. К вычислению коэффициентов уравнения обращенной параболы по методу средних

Температура, °C, x	Прирост, mm, y	$\frac{x}{y}$	$\frac{x_a}{(x-15)}$	$\left(\frac{x}{y} - 0,9375\right)$	$\frac{y_a}{x_a}$	y'	$y' - y$	$(y' - y)^2$
5	5	1	-10	0,0625	-0,00625	3,59	1,41	1,9881
10	11	0,9091	-5	-0,0284	0,00568	11,07	0,07	0,0049
$x_1 = 15$	16	$\left(\frac{x}{y}\right)_1 = 0,9375$	0	0	0	16,73	0,73	0,5329
20	15	1,3333	5	0,3958	0,07916	14,56	0,44	0,1936
25	10	2,5000	10	1,5625	0,15625	10,71	0,71	0,5041
<hr/>								
$\sum_1 x = 75$	—	—	—	—	$\sum_1 \frac{y_a}{x_a} = 0,23484$	—	—	—
<hr/>								
30	8	3,7500	15	2,8125	0,18750	7,94	0,06	0,0036
35	6	5,8333	20	4,8958	0,24479	6,13	0,13	0,0169
40	5	8,0000	25	7,0625	0,28250	4,93	0,07	0,0049
45	4	11,2500	30	10,3125	0,34375	4,09	0,09	0,0081
<hr/>								
$\sum_2 x = 150$	—	—	—	—	$\sum_2 \frac{y_a}{x_a} = 1,05854$	—	—	—
$\sum x = 225$	—	$\sum_1 \frac{x}{y} = 35,5132$	—	—	—	—	—	$\sum x^2 = 7125$
<hr/>								

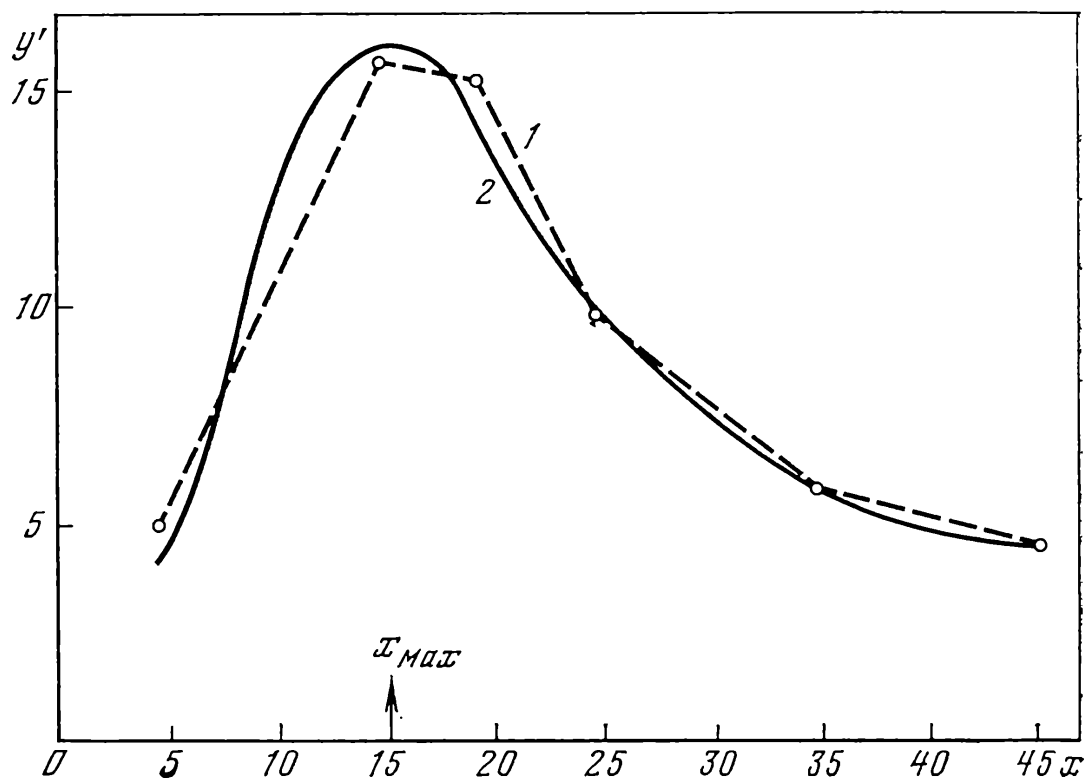


Рис. 20. Зависимость прироста грибницы (y) от температуры среды (x)
1 — эмпирические данные; **2** — теоретическая кривая

Решая эти уравнения с учетом того, что $a'_1 = a_1 + 15a_2$, получаем: $a_2 = 0,0096741$; $a'_1 = -0,098143$; $a_1 = -0,24325$.

5. Величину коэффициента a_0 можно определить при помощи уравнения параболы второго порядка:

$$\Sigma(x/y) = Na_0 + a_1\Sigma x + a_2\Sigma x^2, \quad (3.60)$$

где $\Sigma(x/y)$ — сумма величин столбца 3; N — объем выборки; a_0 , a_1 , a_2 — коэффициенты уравнения (3.58); Σx — сумма величин столбца 1; Σx^2 — сумма квадратов величин столбца 1, в данном случае $\Sigma x^2 = 7125$ (столбец 10).

Подставим суммы из табл. 3.10 в уравнение (3.60): $35,5132 = 9a_0 + (-0,24325) \cdot 225 + 0,0096741 \cdot 7125$; отсюда: $a_0 = 2,3686$.

Итак, зависимость прироста грибницы от температуры окружающей среды может быть аппроксимирована уравнением:

$$y' = x/(2,3686 - 0,24325x + 0,0096741x^2).$$

Для вычисления ошибки данного уравнения найдены теоретические значения функции y' (столбец 7), их разности с эмпирическими данными (столбцы 2 и 8) и квадраты этих разностей (столбец 9). Подставим сумму столбца 9 в формулу (3.04):

$$m_{y \cdot x} = \sqrt{\frac{3,2571}{9-3}} = 0,74 \text{ мм},$$

что и составляет ошибку прогноза прироста грибницы по уравнению обращенной параболы (рис. 20). Абсцисса точки максимума этой

Таблица 3.11. К вычислению коэффициентов обращенной параболы по второму способу

x	y	y_0		x_0	x_0^2	x_0^4	y_0	$y_0 x_0$	$y_0 x_0^2$	y^2
1	2	3	4	5	6	7	8	9	10	11
5	5	1	0	—4	16	256	—2,9459	11,7836	—47,1344	4,35
10	11	0,90909	1	—3	9	81	—3,0368	9,1104	—27,3312	13,1
15	16	0,9375	2	—2	4	16	—3,0084	6,0168	—12,0336	17,8
20	15	1,33333	3	—1	1	1	—2,6126	2,6126	—2,6126	14,4
25	10	2,5	4	0	0	0	—1,4459	0	0	10,4
30	8	3,75	5	1	1	1	—0,1959	—0,1959	—0,1959	7,75
35	6	5,8333	6	2	4	16	1,8874	3,7748	7,5496	6,03
40	5	8	7	3	9	81	4,0541	12,1623	36,4869	4,87
45	4	11,25	8	4	16	256	7,3041	29,2164	116,8656	4,06
<hr/>										
35,51325			36	0	60	708	0,0001	74,481	71,5944	

кривой:

$$x = \sqrt{(a_0/a_2)}, \quad (3.61)$$

где обозначения те же, что и к формуле (3.58). Для рассмотренного примера абсцисса точки максимума:

$$x = \sqrt{(2,3686/0,00967)} = 15,6.$$

Если кривая, найденная методом средних, неудовлетворительно совпадает с эмпирическими данными, следует попытаться улучшить качество аппроксимации путем повторения вычислений с другой парой избранных точек. В рассмотренном случае эти точки были выбраны сравнительно удачно, так как ошибка прогноза невелика.

Второй способ. По способу наименьших квадратов и с преобразованием исходных данных коэффициенты обращенной параболы вычисляются следующим образом.

1. Приведенные в столбцах 1, 2 табл. 3.11 величины аргумента делим на значения функции, и результаты ($y_0 = x/y$) записываем в столбце 3.

2. Для сокращения вычислений преобразуем величины x по формуле: $x' = (x - A)/c$, где x' — преобразованные, а x — исходные величины; A — начальное значение ряда x , который при этом способе должен представлять собой арифметическую прогрессию с разностью членов, равной c . В нашем примере $A = 5$; $c = 5$.

3. Найдем среднюю арифметическую ряда x' : $M_1 = (\Sigma x')/N = 36/9 = 4$. Объем исходных рядов данных $N = 9$.

4. Преобразуем величины x' в x_0 по формуле: $x_0 = x' - M_1$, результаты запишем в столбце 5.

5. В столбцах 6 и 7 записаны квадраты и четвертые степени величин x_0 .

6. Вычислим среднюю арифметическую ряда y_b :

$$M_2 = (\Sigma y_b)/N = 35,51325/9 = 3,9459.$$

7. Преобразуем величину y_b в y_0 по формуле:

$$y_0 = y_b - M_2 \text{ (столбец 8).}$$

С величиной y_0 произведены вычисления, понятные из заголовков столбцов 9, 10.

8. Найдем коэффициенты параболы для величин x_0, y_0 по формулам (3.44)–(3.46):

$$a'_2 = \frac{71,5944}{708 - \frac{60^2}{9}} = 0,23245; \quad a'_1 = \frac{74,481}{60} = 1,24135;$$

$$a'_0 = -\frac{0,23245 \cdot 60}{9} = -1,54967,$$

что дает уравнение: $y_0 = -1,54967 + 1,24135x_0 + 0,23245x_0^2$, по которому с помощью выражения

$$y' = x/(y_0 + M_2) \quad (3.62)$$

можно найти теоретические значения функции уравнения обращенной параболы, где $x_0 = \frac{x-A}{c} - M_1$.

9. Если требуется получить коэффициенты уравнения обращенной параболы в размерности для исходных данных, то далее производим следующие вычисления по формулам:

$$a_0 = a'_0 + M_1(a'_2M_1 - a'_1) + \frac{A}{c} \left[a'_2 \left(2M_1 + \frac{A}{c} \right) - a'_1 \right] + M_2, \quad (3.63)$$

$$a_1 = \frac{a'_1 - 2a'_2 \left(\frac{A}{c} + M_1 \right)}{c}, \quad a_2 = \frac{a'_2}{c^2},$$

$$a_0 = -1,54967 + 4(0,23245 \cdot 4 - 1,24135) + 5/5 [0,23245(2 \cdot 4 + 5/5) - 1,24135] + 3,9459 = 2,0005;$$

$$a_1 = [1,24135 - 2 \cdot 0,23245(5/5 + 4)]/5 = -0,21664;$$

$$a_2 = 0,23245/5^2 = 0,009298,$$

что дает уравнение обращенной параболы для исходных данных:

$$y' = \frac{x}{2,0005 - 0,21664x + 0,009298x^2}.$$

Заметим, что удобнее вычислять теоретические значения функции обращенной параболы по любой из следующих равносильных формул:

$$y = \frac{1}{\frac{a_0}{x} + a_1 + a_2x}, \quad (3.64)$$

$$y = \frac{x}{a_0 + x(a_1 + a_2x)} \quad (3.65)$$

Теоретические значения функции для нашего примера даны в столбце 11 табл. 3.11, их ошибка (3.04): $m_{y \cdot x} = \sqrt{\frac{8,6283}{9-3}} = 1,2$.

§ 3.06. Вычисление коэффициентов уравнений парабол способом сумм

Рассматриваемые в этом параграфе схемы вычисления эффективны при использовании как автоматических, так и ручных средств счета, в последнем случае с применением специальных таблиц. Ряды аргумента могут быть с четным и нечетным числом вариантов, более желателен второй случай, при котором несколько сокращаются расчеты. Варианты ряда аргумента должны располагаться через равные интервалы. Существенно сокращает вычисления предварительное приведение рядов аргумента и функции к малозначным числам (0, ± 1 порядка), что всегда несложно сделать при помощи простейших арифметических преобразований. По приведенным способам сумм можно сравнительно легко рассчитать три уравнения парабол: первого, второго и третьего порядка. При этом благодаря системе поэтапной проверки вычислений в значительной степени гарантируются правильные результаты. Алгоритмы способа сумм для четных и нечетных (по числу вариантов) рядов различаются, поэтому рассмотрим их отдельно на биологических примерах.

а. Ряд с нечетным числом вариантов.

В первых двух столбцах табл. 3.12 приведены числа, отражающие зависимость числа побегов (y'') от числа соцветий (x') на растениях гелениума осеннего. Как это было рекомендовано, приведем указанные величины по возможности ближе к 0, ± 1 порядку, для чего число побегов в столбце 4 выразим в десятках ($y = y''/10$). Число соцветий преобразовано в столбце 3 по формуле:

$$x = \frac{x' - A}{c} = \frac{x' - 486}{134},$$

Таблица 3.12. Вычисление сумм произведений функции (для нечетного N)

Число соцветий, x'	Число побегов, y''		y	y_1	y_2	y_3	y_4
1	2	3	4	5	6	7	8
84	4,1	—3	0,41	0,41	0,41	0,41	0,41
218	6,9	—2	0,69	1,10	1,51	1,92	—
352	9,7	—1	0,97	2,07	3,58	—	—
486	10,1	0	1,01	—	—	—	—
620	12,4	1	1,24	5,28	11,92	21,16	33
754	14,4	2	1,44	4,04	6,64	9,24	11,84
888	26,0	3	2,6	2,6	2,6	2,6	2,6
$c = 134$	$k = 7$	0	8,36				
			p_i	2,07	3,58	1,92	0,41
			q_i	5,28	11,92	21,16	33

где x — преобразованные, а x' — исходные величины аргумента; A — центральное число ряда x' ; c — классовой интервал ряда x' . В соответствии с данным алгоритмом преобразование величин ряда функции желательно, а указанное преобразование ряда аргумента обязательно.

1. Дальнейшие действия в столбцах 5—8 являются четырехкратным суммированием величин y . В таблице необходимо поставить 7 черточек, начиная с центральной, нулевой строки. Затем в верхней и нижней строках повторить крайние варианты ряда функции: 0,41 и 2,6. Суммирование во всех столбцах ведется по направлению от крайних к центральной строке: $0,41 + 0,69 = 1,10$; $1,10 + 0,97 = 2,07$ — суммирование прекращаем, так как в следующей клетке — черточка. В нижней части столбца 5 числа получены так: $2,6 + 1,44 = 4,04$; $4,04 + 1,24 = 5,28$. Так же получены числа в остальных столбцах 6—8.

2. Выписываем внизу таблицы нижнее число (p_i) верхней половины и верхнее число (q_i) нижней половины каждого столбца и вычисляем суммы произведений функции: $\Sigma y = p_1 + q_1 + y_0 = 2,07 + 5,28 + 1,01 = 8,36$. Сравниваем результат с суммой вариантов ряда y , они должны совпасть; y_0 — здесь обозначает центральную варианту ряда y .

$$\Sigma xy = q_2 - p_2 = 11,92 - 3,58 = 8,34;$$

$$\Sigma x^2 y = (p_3 + q_3) 2 - \Sigma xy = (1,92 + 21,16) 2 - 8,34 = 37,82; \quad (3.66)$$

$$\Sigma x^3 y = [(q_4 - p_4) - (p_3 + q_3)] 6 + \Sigma xy = [(33 - 0,41) - (1,92 + 21,16)] \cdot 6 + 8,34 = 65,4.$$

3. Проверяем вычисление сумм функций, для чего повторяем его по иному способу, который можно использовать при желании и как основной.

В табл. 3.13 суммирование производится так же, но в центре таблицы располагаются симметрично 9 черточек. В нижних строках таблицы: суммы верхних (Σ_v) и нижних (Σ_n) половин столбцов, суммы (s) и разности (d) сумм верхних и нижних половин столбцов.

Суммы произведений функции находим по формулам:

$$\Sigma y = s_0 = 8,36; \quad \Sigma xy = d_1 = 8,34;$$

$$\Sigma x^2 y = 2s_2 + s_1 = 2 \cdot 11,16 + 15,5 = 37,82; \quad (3.67)$$

$$\Sigma x^3 y = 6(d_3 + d_2) + d_1 = 6(2,19 + 7,32) + 8,34 = 65,4.$$

Величины сумм совпали с найденными по формулам (3.66), следовательно, их вычисление выполнено правильно.

4. Коэффициенты уравнений парабол первого, второго и третьего порядков находим при помощи табл. 29П, откуда берем величины множителей $M_1 - M_8$ для данного $N = 7$.

Таблица 3.13. Проверка вычисления сумм произведений функции
(для нечетного N)

	y	y_1	y_2	y_3
1	2	3	4	5
—3	0,41	0,41	0,41	0,41
—2	0,69	1,10	1,51	—
—1	0,97	2,07	—	—
0	1,01	—	—	—
1	1,24	5,28	—	—
2	1,44	4,04	6,64	—
3	2,6	2,6	2,6	2,6
\sum_B	—	3,58	1,92	0,41
\sum_H	—	11,92	9,24	2,6
s_i	8,36	15,5	11,16	3,01
d_i	—	8,34	7,32	2,19

Парабола первого порядка (прямая линия):

$$a_0 = M_1 \Sigma y = 0,14286 \cdot 8,36 = 1,1942;$$

$$a_1 = M_2 \Sigma xy = 0,035714 \cdot 8,34 = 0,29785; \quad (3.68)$$

$$y' = 1,1942 + 0,29785x.$$

Парабола второго порядка:

$$a_0 = M_3 \Sigma y - M_4 \Sigma x^2 y = 0,33333 \cdot 8,36 - 0,047619 \cdot 37,82 = 0,98569;$$

$$a_1 = M_2 \Sigma xy = 0,035714 \cdot 8,34 = 0,029785. \quad (3.69)$$

$$a_2 = M_5 \Sigma x^2 y - M_4 \Sigma y = 0,0119048 \cdot 37,82 - 0,047619 \cdot 8,36 = 0,052145;$$

$$y' = 0,98569 + 0,29785x + 0,052145x^2.$$

Парабола третьего порядка:

$$a_0 = M_3 \Sigma y - M_4 \Sigma x^2 y = 0,33333 \cdot 8,36 - 0,047619 \cdot 37,82 = 0,98569;$$

$$a_1 = M_6 \Sigma xy - M_7 \Sigma x^3 y = 0,26257 \cdot 8,34 - 0,032407 \cdot 65,4 = 0,0704; \quad (3.70)$$

$$a_2 = M_5 \Sigma x^2 y - M_4 \Sigma y = 0,0119048 \cdot 37,82 - 0,047619 \cdot 8,36 = 0,052145;$$

$$a_3 = M_8 \Sigma x^3 y - M_7 \Sigma xy = 0,0046296 \cdot 65,4 - 0,032407 \cdot 8,34 = 0,0325.$$

$$y' = 0,98569 + 0,0704x + 0,052145x^2 + 0,0325x^3.$$

Во всех трех уравнениях: $x = 0, \pm 1, \pm 2, \pm 3$.

При отсутствии таблиц или при счете на ЭВМ множители $M_1 - M_8$ для нечетного числа вариантов в ряду можно вычислить по формулам:

$$\begin{aligned} M_1 &= \frac{1}{N}, & M_2 &= \frac{12}{N(N^2-1)}, & M_3 &= \frac{3(3N^2-7)}{4N(N^2-4)}, \\ M_4 &= \frac{15}{N(N^2-4)}, & M_5 &= \frac{180}{N(N^2-1)(N^2-4)}, \\ M_6 &= \frac{25[N^2(N^2-18)+31]}{N(N^2-1)(N^2-4)(N^2-9)}, \\ M_7 &= \frac{140(3N^2-7)}{N(N^2-1)(N^2-4)(N^2-9)}, & M_8 &= \frac{2800}{N(N^2-1)(N^2-4)(N^2-9)}, \end{aligned} \quad (3.71)$$

где N — объем, ряда.

Вычисление этих множителей можно также проверить по формулам:

$$\begin{aligned} M_1 &= \frac{1}{N}, & M_2 &= \frac{1}{\sum x^2}, & M_3 &= \frac{\sum x^4}{D_2}, & M_4 &= \frac{\sum x^2}{D_2}, \\ M_5 &= \frac{N}{D_2}, & M_6 &= \frac{\sum x^6}{D_3}, & M_7 &= \frac{\sum x^4}{D_3}, & M_8 &= \frac{\sum x^2}{D_3}, \end{aligned} \quad (3.72)$$

где D_2, D_3 — определители распределения ряда натуральных чисел, их величины даны в табл. 31П, где они вычислены по формулам:

$$D_2 = \frac{N^2(N-1)(N^2-4)}{180}, \quad D_3 = \frac{N^2(N^2-1)(N^2-4)(N^2-9)}{33\,600}. \quad (3.73)$$

Вычисление определителей можно проверить по формулам:

$$D_2 = N \cdot \sum x^4 - (\sum x^2)^2; \quad D_3 = \sum x^2 \cdot \sum x^6 - (\sum x^4)^2 \quad (3.74)$$

Суммы степеней отклонений натуральных чисел от их средней приведены в табл. 31П или могут быть для нечетного N вычислены по формулам:

$$\begin{aligned} \sum x^2 &= \frac{N(N^2-1)}{12}, & \sum x^4 &= \sum x^2 \cdot \frac{3N^2-7}{20}, \\ \sum x^6 &= \sum x^2 \cdot \frac{3N^2(N^2-6)+31}{112}, \end{aligned} \quad (3.75)$$

а для четного N — по формулам:

$$\begin{aligned}\sum x^2 &= \frac{N(N^2 - 1)}{3}, & \sum x^4 &= \sum x^2 \cdot \frac{3N^2 - 7}{5}, \\ \sum x^6 &= \sum x^2 \cdot \frac{3N^2(N^2 - 6) + 31}{7}.\end{aligned}\quad (3.76)$$

5. Вычисление коэффициентов парабол, приведенное в предыдущем пункте, проверяем с помощью табл. 31П по следующим формулам.

Парабола первого порядка (прямая):

$$\begin{aligned}a_0 &= \frac{\sum y}{N}, & a_1 &= \frac{\sum xy}{\sum x^2}, \\ a_0 &= 8,36/7 = 1,194; & a_1 &= 8,34/28 = 0,2978.\end{aligned}\quad (3.77)$$

Парабола второго порядка:

$$a_0 = \frac{\sum y \cdot \sum x^4 - \sum x^2 y \cdot \sum x^2}{D_2}, \quad a_1 = \frac{\sum xy}{\sum x^2}, \quad (3.78)$$

$$a_2 = \frac{N \sum x^2 y - \sum y \cdot \sum x^2}{D_2},$$

$$a_0 = \frac{8,36 \cdot 196 - 37,82 \cdot 28}{588} = 0,9857;$$

$$a_1 = \frac{8,34}{28} = 0,2978; \quad a_2 = \frac{7 \cdot 37,82 - 8,36 \cdot 28}{588} = 0,052143.$$

Парабола третьего порядка:

$$a_0 = \frac{\sum y \sum x^4 - \sum x^2 y \sum x^2}{D_2}, \quad a_1 = \frac{\sum xy \sum x^6 - \sum x^3 y \sum x^4}{D_3}, \quad (3.79)$$

$$a_2 = \frac{N \sum x^2 y - \sum y \sum x^2}{D_2}, \quad a_3 = \frac{\sum x^2 \sum x^3 y - \sum x^4 \sum xy}{D_3},$$

$$a_0 = \frac{8,36 \cdot 196 - 37,82 \cdot 28}{588} = 0,9857;$$

$$a_1 = \frac{8,34 \cdot 1588 - 65,4 \cdot 196}{6048} = 0,070357;$$

$$a_2 = \frac{7 \cdot 37,82 - 8,36 \cdot 28}{588} = 0,052143;$$

$$a_3 = \frac{28 \cdot 65,4 - 196 \cdot 8,34}{6048} = 0,0325.$$

По найденным уравнениям парабол удобно вычислять теоретические значения функции, представив их в виде: $y' = a_0 + x(a_1 + a_2x)$ — парабола второго порядка, $y' = a_0 + x(a_1 + x(a_2 + a_3x))$ — парабола третьего порядка, где $x = 0, \pm 1, \pm 2, \pm 3$.

Таблица 3.14. Вычисление y' по уравнению параболы третьего порядка

$\alpha_3 x$	(1) + a_2		(3) + a_1		(5) + $a_0 = y'$	x
$0,0325 \cdot x$	(1) + 0,052143	(2) · x	(3) + 0,070357	(4) · x	(5) + 0,98569	
1	2	3	4	5	6	7
—0,0975	—0,04536	+0,13608	+0,20644	—0,61932	0,36637	—3
—0,065	—0,01286	+0,02572	+0,096077	—0,19215	0,79354	—2
—0,0325	+0,01964	—0,01964	+0,050717	—0,050717	0,93497	—1
0	+0,052143	+0	+0,070357	0	0,98569	0
0,0325	+0,08464	+0,0864	+0,156757	0,156757	1,14245	1
0,065	+0,11714	+0,23428	+0,30464	0,60928	1,59497	2
0,0975	+0,14964	+0,44892	+0,519277	1,55783	2,5435	3

В табл. 3.14 приведено вычисление теоретических значений функции кубической параболы, представленной в форме последнего уравнения, для рассматриваемого примера. В столбце 6 записаны теоретические значения числа побегов гелениума, которые хорошо соответствуют их эмпирическим значениям (см. столбец 4 табл. 3.12). Теоретические значения функции, вычисленные по уравнению параболы первого порядка $y' = 1,194 + 0,298x$: 0,3; 0,6; 0,9; 1,2; 1,5; 1,8; 2,1, а по уравнению параболы второго порядка $y' = 0,985 + x (0,298 + 0,052 x)$: 0,56; 0,60; 0,74; 0,98; 1,33; 1,79; 2,35.

б. Ряд с четным числом вариантов.

В столбцах 1 и 3 табл. 3.15 представлены данные, отражающие сезонное изменение среднемесячной температуры воздуха в Ленинграде. Рассчитаем ординаты параболы третьего порядка, аппроксимирующей эту зависимость.

1. При вычислении по данной схеме следует кодировать ряд аргумента так, чтобы его варианты, начиная от центра, были представлены числами: $\pm 0,5$; $\pm 1,5$; $\pm 2,5$ и т. д. Для этого находим условную среднюю (A) и величину интервала (c), составляя два уравнения по двум центральным вариантам:

$$-0,5 = \frac{6 - A}{c}, \quad +0,5 = \frac{7 - A}{c},$$

откуда $A = 6,5$; $c = 1$, после чего по формуле

$$x = \frac{x' - A}{c} = \frac{x' - 6,5}{1}$$

кодируем варианты аргумента (столбец 2).

2. По тому же принципу, что и в табл. 3.13, суммируем числа в столбцах 4, 5, 6, располагая в них 6 черточек указанным способом около центральной линии.

3. Находим суммы верхней (Σ_v) и нижней (Σ_n) половины чисел в тех же столбцах, затем их суммы и разности, причем вычитается верхняя сумма из нижней: $\Sigma_n - \Sigma_v$.

Таблица 3.15. Суммы произведений функции для вычисления кубической параболы с четным числом вариант в ряду

Месяц, x'		Средняя температура, °С. y	y_1	y_2	y_3	y'	y''
1	2	3	4	5	6	7	8
1	—5,5	—6,5	—6,5	—6,5	—6,5	—9,7	—12,8
2	—4,5	—7	—13,5	—20	—26,5	—5,0	—4,7
3	—3,5	—4,4	—17,9	—37,9	—64,4	—0,1	1,8
4	—2,5	3,6	—14,3	—52,2	—116,6	4,6	7
5	—1,5	9,8	—4,5	—56,7	—	8,9	10,6
6	—0,5	15,3	10,8	—	—	12,2	12,8
7	0,5	17,5	46,0	—	—	14,2	13,6
8	1,5	16	28,5	33,8	—	14,6	12,9
9	2,5	11,2	12,5	5,3	—14,5	13,1	10,7
10	3,5	5,7	1,3	—7,2	—19,8	9,1	7,1
11	4,5	—0,3	—4,4	—8,5	—12,6	2,3	2,1
12	5,5	—4,1	—4,1	—4,1	—4,1	—7,5	—4,4
$\sum_{\text{В}}$		10,8	—45,9	—173,3	—214		
$\sum_{\text{Н}}$		46	79,8	19,3	—51		
s		56,8	—	—154	—		
d		35,2	125,7	192,6	163		

4. Находим искомые суммы произведений функции по формулам:

$$\sum y = s_0 = 56,8;$$

$$\sum xy = d_1 - \frac{d_0}{2} = 125,7 - \frac{35,2}{2} = 108,1;$$

$$\sum x^2y = 2s_2 + \frac{s_0}{4} = 2 \cdot (-154) + \frac{56,8}{4} = -293,8;$$

$$\begin{aligned} \sum x^3y &= 3(2d_3 + d_2) + \frac{\sum xy}{4} = 3(2 \cdot 163 + 192,6) + \\ &+ \frac{125,7 - 0,5 \cdot 35,2}{4} = 1582,825. \end{aligned} \quad (3.80)$$

5. При ручном счете целесообразно далее ввести проверочный расчет, при котором варианты аргумента кодируются, как показано в табл. 3.16, для чего по двум центральным вариантам составляем уравнения: $-1 = \frac{6-A}{c}$, $1 = \frac{7-A}{c}$, откуда $A = 6,5$; $c = 0,5$.

По формуле:

$$x = \frac{x' - A}{c} = \frac{x' - 6,5}{0,5},$$

Таблица 3.16. Проверка вычисления сумм произведений функции для уравнения кубической параболы (ряд с четным числом вариант)

Месяц, x'		Средняя температура, °C, y	y_1	y_2	y_3	y_4	y'
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	—11	—6,5	—6,5	—6,5	—6,5	—6,5	—9,7
2	—9	—7	—13,5	—20	—26,5	—33	—5,0
3	—7	—4,4	—17,9	—37,9	—64,4	—97,4	—0,1
4	—5	3,6	—14,3	—52,2	—116,6	—214	4,6
5	—3	9,8	—4,5	—56,7	—173,3	—	8,9
6	—1	15,3	10,8	—	—	—	12,2
7	+1	17,5	46	79,8	—	—	14,2
8	+3	16,0	28,5	33,8	19,3	—31,7	14,6
9	+5	11,2	12,5	5,3	—14,5	—51	13,1
10	+7	5,7	1,3	—7,2	—19,8	—36,5	9,1
11	+9	—0,3	—4,4	—8,5	—12,6	—16,7	2,3
12	+11	—4,1	—4,1	—4,1	—4,1	—4,1	—7,5
p_i			10,8	—56,7	—173,3	—214	
q_i			46	79,8	19,3	—31,7	

вычислим кодированные величины вариант аргумента и запишем их в столбец 2. Суммирование чисел в столбцах 4—7 ведется так же, как и в предыдущей таблице. Черточки около центральной черты необходимо располагать так, как указано в табл. 3.16. Числа над центральной чертой (p) и непосредственно под ней (q) выписываем снизу таблицы. Суммы произведений находим по формулам:

$$\begin{aligned}\Sigma'y &= p_1 + q_1 = 10,8 + 46 = 56,8; \\ \Sigma'xy &= 2(q_2 - p_2) - \Sigma y = 2(79,8 + 56,7) - 56,8 = 216,2;\end{aligned}\tag{3.81}$$

$$\begin{aligned}\Sigma'x^2y &= 8(p_3 + q_3) + \Sigma y = 8(19,3 - 173,3) + 56,8 = -1175,2; \\ \Sigma'x^3y &= 24[2(q_4 - p_4) - (p_3 + q_3)] + \Sigma xy = 24[2(-31,7 + \\ &+ 214) - (-173,3 + 19,3)] + 216,2 = 12662,6.\end{aligned}$$

Приводимый в пункте 5 алгоритм вычисления сумм может быть использован не только для проверки, но и в качестве самостоятельного; в этом случае дальше следует использовать суммы в том виде, как они только что получены по формулам (3.81). С целью проверки эти суммы надо соответственно разделить на $2^0, 2^1, 2^2, 2^3$, т. е.:

$$\begin{aligned}\Sigma y &= \Sigma'y; \Sigma xy = (\Sigma'xy)/2 = 216,2/2 = 108,1; \\ \Sigma x^2y &= (\Sigma'x^2y)/4 = -1175,2/4 = -293,8; \\ \Sigma x^3y &= (\Sigma'x^3y)/8 = 12662,6/8 = 1582,825.\end{aligned}\tag{3.82}$$

6. По формулам (3.79), с применением табл. 31П вычисляем коэффициенты уравнения кубической параболы:

$$a_0 = \frac{56,8 \cdot 3038,75 + 293,8 \cdot 143}{16016} = 13,4;$$

$$a_1 = \frac{108,1 \cdot 76156,4375 - 1582,825 \cdot 3038,75}{1656369} = 2,066379;$$

$$a_2 = \frac{-12 \cdot 293,8 - 56,8 \cdot 143}{16016} = -0,72727;$$

$$a_3 = \frac{143 \cdot 1582,825 - 3038,75 \cdot 108,1}{1656369} = -0,061668.$$

Вычисление коэффициентов можно проверить по формулам (3.70) с применением таблицы множителей для четных значений N (табл. 30П):

$$a_0 = 0,1897321 \cdot 56,8 - 0,0089986 \cdot (-293,8) = 13,42057;$$

$$a_1 = 0,0459779 \cdot 108,1 - 0,0018346 \cdot 1582,825 = 2,06636;$$

$$a_2 = 0,0007493 \cdot (-293,8) - 0,0089986 \cdot 56,8 = -0,73126;$$

$$a_3 = 0,0000863 \cdot 1582,825 - 0,0018346 \cdot 108,1 = -0,061722.$$

Из-за приближенных значений множителей, приведенных в табл. 30П, в величинах коэффициентов наблюдаются расхождения. В данном случае эти неточности почти не повлияли на величины ординат параболы, но при необходимости можно уточнить значения множителей по формулам (3.71), (3.72).

7. По уравнению:

$y' = 13,4 + x [2,066379 + x (-0,72727 - 0,061668x)]$, где $x = \pm 0,5; \pm 1,5; \pm 2,5$. или $x = (x' - 6,5)/1,0$, вычисляем ординаты кубической параболы, которые записаны в столбце 7 табл. 3.15.

Малая величина коэффициента a_3 в полученном уравнении кубической параболы дает основание полагать, что данная эмпирическая зависимость может быть приближенно аппроксимирована также параболой второго порядка, коэффициенты уравнения которой находим по формулам (3.78)

$$a_0 = (56,8 \cdot 3038,75 + 293,8 \cdot 143)/16016 = 13,4;$$

$$a_1 = 108,1/143 = 0,75594;$$

$$a_2 = (12 \cdot (-293,8) - 56,8 \cdot 143)/16016 = -0,72727.$$

Уравнение этой параболы: $y'' = 13,4 + 0,75594x - 0,72727x^2$, где $x = \pm 0,5; \pm 1,5; \pm 2,5$. ; его ошибка равна по формуле (3.04)

$$m_{y \cdot x} = \sqrt{\frac{134,28}{12 - 3}} = 3,86.$$

Ординаты параболы второго порядка (y'') вычислены по приведенному уравнению и записаны (округленно) в столбце 8 табл. 3.15.

Уравнение прямой для данного случая может быть вычислено по формулам (3.77).

8. Рассчитаем также коэффициенты уравнения и ординаты кубической параболы по суммам, найденным в табл. 3.16. По формулам (3.70) и с множителями из табл. 29П, коэффициенты равны:

$$a_0 = 0,189732 \cdot 56,8 - 0,00223214 \cdot (-1175,2) = 13,4;$$

$$a_1 = 0,0114945 \cdot 216,2 - 0,000114662 \cdot 12662,6 = 1,033192;$$

$$a_2 = 0,0000468282 \cdot (-1175,2) - 0,00223214 \cdot 56,8 = -0,181818;$$

$$a_3 = 0,00000134896 \cdot 12662,6 - 0,000114662 \cdot 216,2 = -0,00770858.$$

Ординаты по уравнению:

$$y' = 13,4 + x [1,033192 + x (-0,181818 - 0,00770858x)],$$

где $x = \pm 1; \pm 3; \pm 5$ или $x = (x' - 6,5)/0,5$, записаны в столбце (8) табл. 3.16.

Множители $M_1 - M_8$ для четного числа вариантов в табл. 29П вычислены по формулам:

$$\begin{aligned} M_1 &= \frac{1}{N}, & M_2 &= \frac{3}{N(N^2 - 1)}, & M_3 &= \frac{3(3N^2 - 7)}{4N(N^2 - 4)}, \\ M_4 &= \frac{15}{4N(N^2 - 4)}, & M_5 &= \frac{45}{4N(N^2 - 1)(N^2 - 4)}, \\ M_6 &= \frac{25(3N^4 - 18N^2 + 31)}{4N(N^2 - 1)(N^2 - 4)(N^2 - 9)}, \\ M_7 &= \frac{35(3N^2 - 7)}{4N(N^2 - 1)(N^2 - 4)(N^2 - 9)}, & M_8 &= \frac{175}{4N(N^2 - 1)(N^2 - 4)(N^2 - 9)}. \end{aligned} \quad (3.83)$$

В работе Н. С. Четверикова (1975) содержатся таблицы величин множителей M_{1-8} до $N = 100$ с точностью до 6 знаков мантиссы, а в книге А. К. Митропольского (1971) приводится алгоритм вычисления способом сумм коэффициентов уравнения парабол до 4 порядка включительно.

§ 3.07. Степенная функция

В биологии степенная, или аллометрическая, функция наиболее часто применяется при исследовании морфогенеза и корреляций органов растений и животных; ее уравнение:

$$y' = a_0 x^\beta, \quad (3.84)$$

где y — функция, например величина одного из органов; a_0 — коэффициент, соответствующий константе начального роста второго органа или особи в целом; x — аргумент, или величина, второго органа или особи, измеренная в тот же момент времени; β — коэффициент, или константа аллометрии, показывающая во сколько раз скорость роста первого органа (y) превышает скорость роста второго органа (x). Если $\beta = 1$, оба органа растут с одинаковой скоростью и

Таблица 3.17. К вычислению параметров уравнения степенной функции

Длина листа, см, x	Площадь листа, см ² , y	$\lg x$	$\lg y$	$(\lg x)^2$	$\lg x \cdot \lg y$	y'	$\delta =$ $y - y'$	δ^2
8	67,65	0,9031	1,8303	0,8156	1,6529	69,92	2,27	5,1529
10	110,60	1,0000	2,0438	1,0000	2,0438	109,32	1,28	1,6384
12	153,40	1,0792	2,1858	1,1647	2,3589	157,51	4,11	16,8921
14	224,10	1,1461	2,3504	1,3135	2,6938	214,49	9,61	92,3521
16	297,50	1,2041	2,4735	1,4499	2,9783	280,22	17,28	298,5984
18	369,15	1,2553	2,5672	1,5758	3,2226	354,89	14,26	203,3476
20	460,40	1,3010	2,6132	1,6926	3,3998	438,13	22,27	495,9529
22	514,50	1,3424	2,7114	1,8020	3,6398	530,40	15,90	252,8100
2197,30		9,2312	18,7756	10,8141	21,9899	$\sum \delta^2 = 1366,7444$		

находятся друг с другом в состоянии изометрии. При $\beta > 1$, скорость роста первого органа (y) больше скорости второго органа (x), что называется положительной аллометрией. При $\beta < 1$ скорость роста второго органа (x) больше скорости первого органа (y) и органы находятся между собой в состоянии отрицательной аллометрии. Таким образом, константа β передает относительную скорость органов в онтогенезе, ее величина приблизительно постоянна для определенных органов и фаз онтогенеза живых существ.

Рассмотрим процесс вычисления уравнения степенной или аллометрической функции на примере аппроксимации зависимости площади листа от его длины у огурца. Вычисления, приведенные в табл. 3.17, выполняются в следующем порядке.

1. Логарифмы величин x и y записаны в столбцах 3 и 4.
2. Логарифмы вариант x (столбец 3) возводим в квадрат (столбец 5).
3. Перемножаем логарифмы x и y (столбцы 3, 4, 6).
4. Суммируем числа в столбцах 2—6.
5. Для нахождения коэффициентов уравнения (3.84), решаем систему уравнений:

$$\begin{aligned} N \lg a_0 + \beta \sum \lg x &= \sum \lg y, \\ \lg a_0 \sum \lg x + \beta \sum (\lg x)^2 &= \sum (\lg x \lg y), \end{aligned} \quad (3.85)$$

где N — число точек или значений эмпирического ряда; a_0 , β — искомые коэффициенты уравнения; прочие обозначения — суммы из табл. 3.17, подставляя которые в нормальные уравнения (3.85), получим: $8 \lg a_0 + 9,2312\beta = 18,7756$, $9,2312 \lg a_0 + 10,8141\beta = 21,9899$, откуда: $\beta = 2,0018$; $a_0 = 1,087$; $\lg a_0 = 0,037026$ и $y' = 1,087 \cdot x^{2,0018}$, или в логарифмической форме: $\lg y' = 0,037026 + 2,0018 \lg x$.

Решить систему уравнений (3.85) можно также по формулам

$$\beta = (A - M_y)/(B - M_x), \quad \lg a_0 = M_y - \beta M_x \quad (3.86)$$

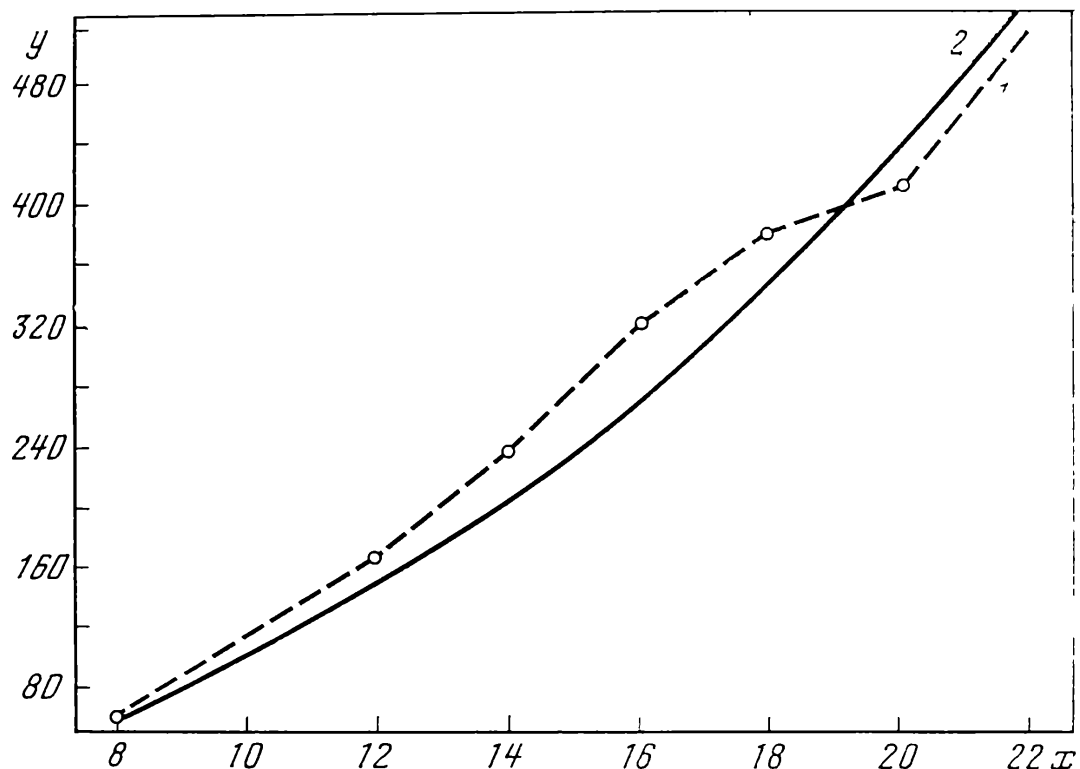


Рис. 21. Зависимость площади листа (y) от его длины (x) у огурца
1 — эмпирическая линия регрессии; 2 — теоретическая кривая степенной функции

и для проверки:

$$\lg a_0 = A - \beta B, \quad (3.87)$$

где M_y — средняя арифметическая логарифмов значений функции из 4 столбца табл. 3.17; M_x — средняя арифметическая логарифмов значений аргумента из 3 столбца табл. 3.17.

$$A = \frac{\sum (\lg x \cdot \lg y)}{\sum \lg x}, \quad B = \frac{\sum (\lg x)^2}{\sum \lg x}, \quad (3.88)$$

$$M_x = 9,2312/8 = 1,1539; \quad M_y = 18,776/8 = 2,347;$$

$$A = 21,99/9,2312 = 2,3821; \quad B = 10,814/9,2312 = 1,1715.$$

Подставив эти величины в формулы (3.86), (3.87), получим те же значения коэффициентов β и a_0 .

В столбцах 7—9 табл. 3.17 приведены теоретические значения площади листа, их разности с фактическими и квадраты разностей, что необходимо для вычисления ошибки полученного уравнения по формуле (3.04):

$$m_{y \cdot x} = \sqrt{\frac{1366,7444}{8-2}} = 15,09 \text{ см}^2.$$

Как видно из рис. 21, фактические данные удовлетворительно совпадают с теоретической линией регрессии. Константа β в 2 раза больше единицы, следовательно, площадь листа возрастает в 2 раза быстрее, чем его длина у огурца и эти два размера листа находятся в отношении положительной аллометрии.

Коэффициенты аллометрического уравнения можно приближенно найти также способом избранных точек по формулам:

$$\beta = \left(\ln \frac{y_1}{y_2} \right) / \left(\ln \frac{x_1}{x_2} \right); \quad (3.89)$$

$$a_0 = \exp \left(\frac{\ln x_2 \ln y_1 - \ln x_1 \ln y_2}{\ln (x_2/x_1)} \right), \quad (3.90)$$

или

$$a_0 = \frac{y_1}{x_1^\beta}, \quad (3.91)$$

где a_0 , β — коэффициенты уравнения; x_1 , y_1 , x_2 , y_2 — координаты двух избранных точек.

Например, для точек с координатами $x_1 = 10$; $y_1 = 110,6$; $x_2 = 20$; $y_2 = 460,4$:

$$\beta = \frac{\ln \frac{110,6}{460,4}}{\ln (10/20)} = 2,0576;$$

$$a_0 = \exp \left(\frac{\ln 20 \cdot \ln 110,6 - \ln 10 \cdot \ln 460,4}{\ln (20/10)} \right) = 0,96448$$

или

$$a_0 = \frac{110,6}{10^{2,0576}} = 0,96661.$$

Более точное приближение к эмпирическим данным дало вычисление функции при a_0 , полученном по формуле (3.90): 69,6; 110,4; 160,5; 220,2; 289,8; 369,1; 458,5; 557,9 (для соответствующих значений x из столбца 1 табл. 3.17).

§ 3.08. Показательно-степенная функция

Эта функция ценна тем, что позволяет аппроксимировать зависимости, которые графически передаются двускатной кривой, в чем в биологических исследованиях необходимость возникает нередко, однако выбор подходящих для этого кривых весьма ограничен. Уравнение показательно-степенной функции:

$$y = ax^b e^{cx} \quad (3.92)$$

или

$$y = \exp (\ln a + b \ln x + cx),$$

где y — функция; x — аргумент; a , b , c — коэффициенты; e — основание натуральных логарифмов.

Коэффициенты этого уравнения можно найти двумя способами: наименьших квадратов и избранных точек. По способу наименьших квадратов требуется выполнить следующие вычисления.

1. Найти средние: $M_x = (\Sigma x)/N$; $M_1 = (\Sigma \ln x)/N$; $M_2 = (\Sigma \ln y)/N$.

Таблица 3.18. Вычисление теоретических значений показательно-степенной функции

x	y	y'		y	y'		y	y'
0,1	1,781	1,029	0,5	1,768	1,829	0,9	0,228	0,22
0,2	3,188	2,556	0,6	1,183	1,183	1,0	0,126	0,116
0,3	3,193	2,975	0,7	0,691	0,71	1,1	0,068	0,06
0,4	2,54	2,54	0,8	0,404	0,404	1,2	0,036	0,03

2. Определить промежуточные величины:

$$A = M_1 \sum \ln x \{ \sum x [\sum \ln y \ln x \cdot (N + 1) - \sum \ln y \sum \ln x] + \\ + \sum x \ln x \cdot \sum \ln y - \sum x \ln y \cdot \sum \ln x \} + \sum \ln^2 x (\sum x \ln y \cdot \sum \ln x - \\ - \sum \ln y \cdot \ln x \cdot \sum x) - \sum x \ln x \cdot \sum \ln x \cdot \sum \ln y \ln x;$$

$$B = M_1 (\sum x \sum \ln x - N \sum x \ln x) (\sum x \ln x - \sum x \sum \ln x + \\ + (\sum x \ln x \cdot \sum x - \sum x^2 \sum \ln x) (M_1 \sum \ln x - \sum \ln^2 x).$$

3. Коэффициенты равны:

$$c = A/B; \quad a = \exp (M_2 - M_1 b - M_x c);$$

$$b = \frac{M_1 \sum \ln y - \sum \ln y \ln x - c (M_x \sum \ln x - \sum x \ln x)}{M_1 \sum \ln x - \sum \ln^2 x}, \quad (3.93)$$

Приведенный алгоритм целесообразно выполнять лишь при помощи ЭВМ с достаточным числом знаков в мантиссе, поэтому рассмотрим числовой пример расчета коэффициентов по более простому способу избранных точек, который также может обеспечить приемлемую точность, особенно при удачном подборе этих точек. В столбцах 1, 2 табл. 3.18 приведены данные К. А. Семендяева (1933), а в столбце 3 — вычисленные по способу избранных точек теоретические значения функции, которые находим в следующем порядке.

1. Из рядов x, y выбираем 3 пары значений аргумента и функции: $x_1 = 0,4; y_1 = 2,54; x_2 = 0,6; y_2 = 1,183; x_3 = 0,8; y_3 = 0,404$. Число вариантов в рядах: $N = 12$.

2. Формулы вычисления коэффициентов:

$$c = \frac{\ln \frac{y_1}{y_2} \cdot \ln \frac{x_2}{x_3} + \ln \frac{y_2}{y_3} \cdot \ln \frac{x_2}{x_1}}{(x_1 - x_2) \ln \frac{x_2}{x_3} + (x_2 - x_3) \ln \frac{x_2}{x_1}}, \quad (3.94)$$

$$b = \frac{\ln \frac{y_2}{y_3} - c (x_2 - x_3)}{\ln \frac{x_2}{x_3}},$$

$$a = \exp (\ln y_1 - b \ln x_1 - c x_1),$$

откуда:

$$c = \frac{\ln \frac{2,54}{1,183} \ln \frac{0,6}{0,8} + \ln \frac{1,183}{0,404} \cdot \ln \frac{0,6}{0,4}}{(0,4 - 0,6) \ln (0,6/0,8) + (0,6 - 0,8) \ln (0,6/0,4)} = -9,16126;$$

$$b = \frac{\ln \frac{1,183}{0,404} + 9,16126 (0,6 - 0,8)}{\ln (0,6/0,8)} = 2,63436;$$

$$a = \exp (\ln 2,54 - 2,63436 \ln 0,4 + 9,16126 \cdot 0,4) = 1108,225.$$

3. По уравнению: $y' = 1108,225x^{2,63436}e^{-9,16126x}$, или $y' = \exp (\ln 1108,225 + 2,63436 \ln x - 9,16126x)$ находим теоретические значения функции (столбец 3).

Ошибка уравнения (3.04):

$$m_{y \cdot x} = \sqrt{\frac{\sum (y - y')^2}{N - n}} = \sqrt{\frac{1,01045}{12 - 3}} = 0,335.$$

Координаты максимального значения функции:

$$x_{max} = -(b/c) = -(2,63436/-9,16126) = 0,28755; \quad (3.95)$$

$$y_{max} = \exp (\ln a + b \ln x_{max} + cx_{max}) = \exp (\ln 1108,225 + 2,63436 \ln 0,28755 - 9,16126 \cdot 0,28755) = 2,9827.$$

В § 3.14 показаны примеры кривых, соответствующих уравнению (3.92). Из ботанических объектов в соответствии с показатель-но-степенной функцией происходит, например, хронологическое изменение числа видов травянистых многолетников, у которых начинается весеннее отрастание побегов.

§ 3.09. Показательная функция

При аппроксимации таких зависимостей, когда величины одного из признаков изменяются по геометрической прогрессии при равномерном изменении величин связанного с ним признака, нередко бывает подходящей показательная функция, уравнение которой

$$y = ae^{bx}, \quad (3.96)$$

где y — функция; x — аргумент; a , b — коэффициенты; e — основание натуральных логарифмов.

Рассмотрим технику расчетов коэффициентов уравнения (3.96) на примере аппроксимации зависимости между датами созревания семян (x) и начала цветения (y) у травянистых многолетников. В результате обработки данных многолетних наблюдений за 1384 видами выявилась эмпирическая зависимость между этими датами, представленная в табл. 3.19. В столбце 1 таблицы приведены даты созревания семян многолетников, а в столбце 2 — соответствующие даты начала цветения в том же массиве видов. Даты даны в днях от 1 марта, перевод в которые и обратно производится по табл. 24П.

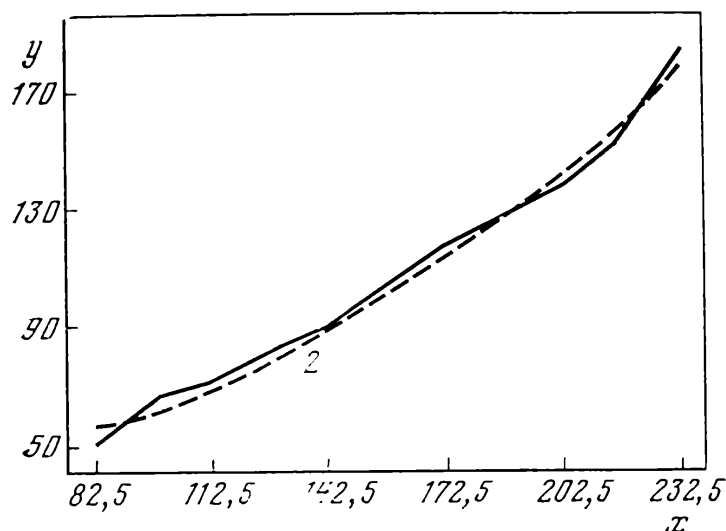


Рис. 22. Зависимость дат начала цветения (y) от дат созревания семян (x) у травянистых многолетников в Москве

1 — эмпирическая, 2 — теоретическая линии регрессии

Таблица 3.19. Даты созревания семян (x) и начала цветения (y) у травянистых многолетников в Москве (в днях от 1 марта)

Созревание семян,	Начало цветения	
	y	y'
82,5	50,0	55,3
97,5	65,5	62,3
112,5	71,3	70,1
127,5	81,0	78,8
142,5	90,0	88,7
157,5	102,7	99,8
172,5	115,3	112,3
187,5	125,5	126,4
202,5	137,3	142,2
217,5	153,5	160,0
232,5	183,1	180,1
$N = 11$		

Коэффициенты вычисляем по формулам:

$$M_1 = \frac{\sum \ln y}{N}, \quad M = \frac{\sum x}{N},$$

$$b = \frac{(\sum x \ln y) / \sum x - M_1}{\sum x^2 / \sum x - M}, \quad a = \exp(M_1 - bM).$$
(3.97)

где M_1 — средняя арифметическая натуральных логарифмов величин функции; $\sum \ln y$ — сумма натуральных логарифмов величин функции; N — объем ряда; M — средняя арифметическая величин аргумента; $\sum x$ — сумма величин аргумента; a, b — искомые коэффициенты; $\sum x \ln y$ — сумма произведений величин аргумента на натуральные логарифмы величин функции; $\sum x^2$ — сумма квадратов величин аргумента.

Вычислив требуемые суммы, подставим их в формулы (3.97) и получим:

$$M_1 = 50,636/11 = 4,6033; \quad M = 1732,5/11 = 157,5;$$

$$b = \frac{8170,0/1732,5 - 4,6033}{297630/1732,5 - 157,50} = 0,0078656;$$

$$a = \exp(4,6033 - 0,0078656 \cdot 157,50) = 28,923.$$

Аппроксимирующее уравнение показательной функции получено:

$$y' = 28,923e^{0,0078656x};$$

вычисленные по нему теоретические значения функции приведены в столбце 3 табл. 3.19. Качество аппроксимации можно считать

вполне удовлетворительным, как это видно из рис. 22, по которому можно заранее прогнозировать даты сбора семян травянистых многолетников или восстанавливать даты начала их цветения, если известны лишь даты созревания семян.

Заметим, что при знаке минус у коэффициента b показательная функция передает зависимость, близкую к гиперболической; в этом случае коэффициент корреляции между признаками также со знаком минус.

§ 3.10. Показательное уравнение второго порядка

При подборе подходящей теоретической линии регрессии к некоторой заданной эмпирической зависимости в настоящее время на практике применяется метод проб и ошибок, так как этот процесс математически пока не формализован. Применение в качестве аппроксимирующей кривой параболы и дальнейшего повышения ее порядка как способа улучшения аппроксимации и отчасти как метода формализации последней ограничивается тем, что далеко не исчерпывает всех видов эмпирических зависимостей, и тем, что повышение порядка параболы выше четвертого обычно нецелесообразно.

В связи с этим представляют интерес различные виды других функций, в том числе показательных, как довольно гибких по своим возможностям. В зависимости от своего строения показательные уравнения принимают вид: $y = 1$ (при $x = 0$); $y = a^x$, $y = a^{bx}$, что можно назвать показательными уравнениями соответственно нулевого, первого и второго порядка. Чем выше его порядок, тем большими возможностями для аппроксимации обладает показательное уравнение. Однако при нахождении коэффициентов показательных уравнений второго порядка возникают значительные трудности, так как методом наименьших квадратов и методом моментов найти коэффициенты не удастся. Метод средних — единственный, применяемый с этой целью, довольно громоздкий и не дает достаточной точности. Итерационные методы в принципе, видимо, применимы в данном случае, не могут быть эффективно реализованы без ЭВМ. Представляется более целесообразным искать коэффициенты рассматриваемых уравнений методом избранных точек, который не уступая по точности методу средних, в то же время значительно проще по технике вычислений.

Основная трудность, возникающая при отыскании коэффициентов показательных уравнений второго порядка, например

$$y = ae^{be^{cx}}, \quad (3.98)$$

заключается в том, что приходится встречаться с решением степенных уравнений вида:

$$k(m^{x_2} - m^{x_3}) = m^{x_1} - m^{x_2}, \quad (3.99)$$

откуда требуется найти m при заданных k , x_1 , x_2 , x_3 . В результате специального исследования удалось найти, что при $x_1 > x_2 > x_3$,

составляющих арифметическую прогрессию:

$$m = k \frac{1}{x_2 - x_3} \quad (3.100)$$

После того, как был найден способ решения степенных уравнений типа (3.99), отыскание коэффициентов уравнения (3.98) стало сравнительно несложной задачей. Логарифмируем это уравнение и напишем его для трех избранных точек: $x_1 > x_2 > x_3$, находящихся на равных интервалах между собой:

$$y_1 = \ln a + be^{cx_1}, \quad y_2 = \ln a + be^{cx_2}, \quad y_3 = \ln a + be^{cx_3}. \quad (3.101)$$

Складывая первое уравнение со вторым и второе с третьим (3.101), получим:

$$\ln y_1 - \ln y_2 = be^{cx_1} - be^{cx_2}, \quad \ln y_2 - \ln y_3 = be^{cx_2} - be^{cx_3}. \quad (3.102)$$

Логарифмируем оба уравнения (3.102), вычитаем второе из первого, затем потенцируем разницу и получаем:

$$\frac{\ln y_1 - \ln y_2}{\ln y_2 - \ln y_3} = \frac{e^{cx_1} - e^{cx_2}}{e^{cx_2} - e^{cx_3}}$$

Обозначая:

$$\frac{\ln \frac{y_1}{y_2}}{\ln \frac{y_2}{y_3}} = k,$$

последнее уравнение запишем в следующем виде:

$$ke^{cx_2} + e^{cx_3} = ke^{cx_3} + e^{cx_1},$$

которое при введении обозначения $m = e^c$ приводится к уравнению (3.99). Подставляя это обозначение в (3.100), получим

$$c = \ln k / (x_2 - x_3),$$

или в более общей форме:

$$c = \frac{\ln k}{|x_i - x_{i+1}|}. \quad (3.103)$$

При известном c из первого уравнения (3.102) найдем:

$$b = \frac{\ln \frac{y_1}{y_2}}{e^{cx_1} - e^{cx_2}}, \quad (3.104)$$

а из первого уравнения (3.101):

$$a = \frac{y_1}{e^{be^{cx_1}}}. \quad (3.105)$$

Пример. Найдем коэффициенты уравнения (3.98) для аппроксимации им следующей эмпирической зависимости:

x	0	1	2	3	4	5	6	7	8	9	10
y	0	0,5	1,6	3,2	5	6,6	7,8	8,6	9,2	9,5	9,6

где x — аргумент; y — функция.

Примем за избранные точки:

$$x_1 = 10; \quad x_2 = 6; \quad x_3 = 2;$$

$$y_1 = 9,6; \quad y_2 = 7,8; \quad y_3 = 1,6.$$

По формулам (3.103)–(3.105):

$$c = \left(\frac{\ln \frac{9,6}{7,8}}{\ln \frac{7,8}{1,6}} \right) / (6 - 2) =$$

$$= -0,50785;$$

$$b = \frac{\ln \frac{9,6}{7,8}}{e^{c10} - e^6} = -5,0312;$$

$$a = \frac{9,6}{e^{bc10}} = 9,9062.$$

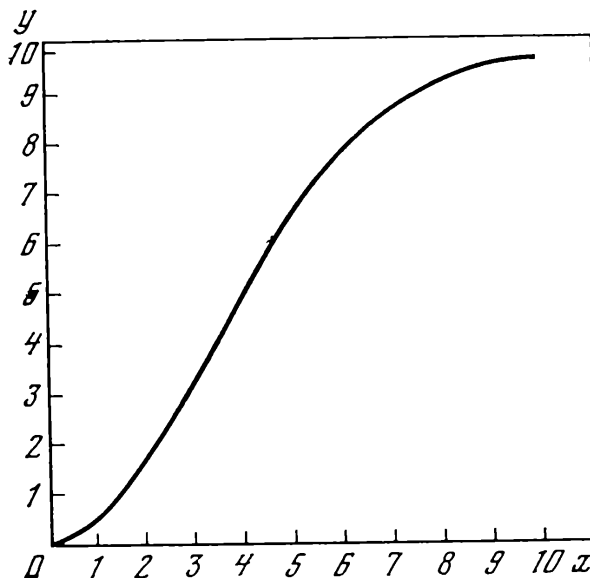


Рис. 23. График показательного уравнения второго порядка

Таким образом, искомое уравнение теоретической линии регрессии предстанет в виде:

$$y = 9,9 e^{-5,03e^{-0,503x}},$$

коэффициенты которого почти совпадают с теми, которые были получены методом средних К. А. Семендяевым (1933, с. 85) для этих же данных.

Форма кривой на рис. 23 указывает, что уравнением (3.98) можно аппроксимировать такие эмпирические зависимости, которые обычно приходится выравнивать более сложными при расчете логистическими кривыми.

§ 3.11. Логистическая функция

Многие явления роста, увеличения численности популяций, изменений физиологического состояния организмов под воздействием неблагоприятных факторов и другие, могут быть моделированы логистической кривой, уравнение которой

$$y' = \frac{a_1}{1 + 10^{\gamma + \beta x}} + a_0, \quad (3.106)$$

где y' — функция, или зависимая переменная; a_0 — нижняя асимптота, или нижний предел функции; a_1 — расстояние между верхней ($a_1 + a_0$) и нижней (a_0) асимптотами (см. рис. 24, А, В); γ , β — коэффициенты уравнения, определяющие изгиб и наклон логистической кривой.

Рассмотрим процесс вычисления значений асимптот и коэффициентов уравнения (3.106) на примере аппроксимации динамики накопления сухого вещества в растущих плодах томата.

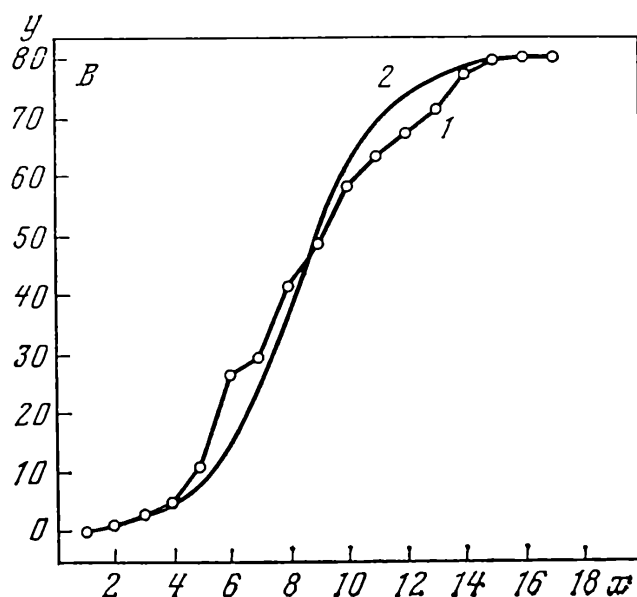
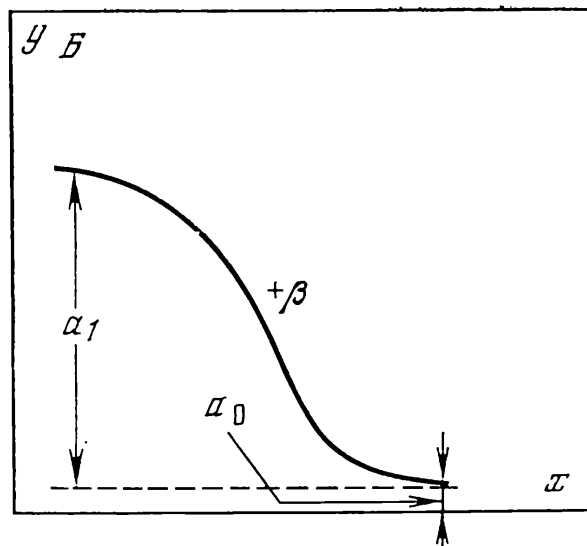
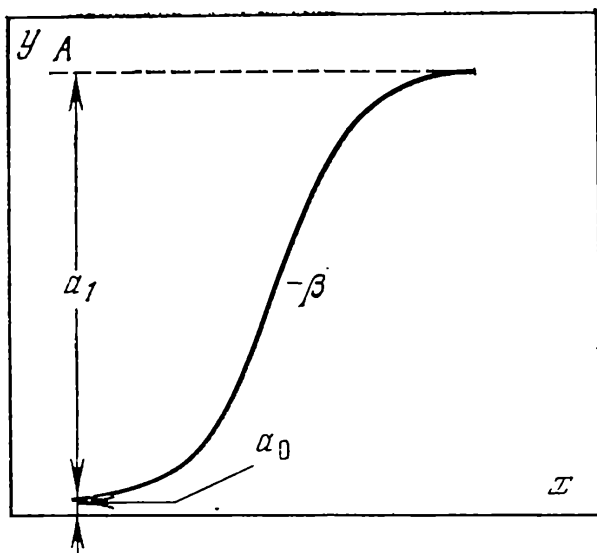


Рис. 24. Логистическая функция

А — прямая,
Б — обратная зависимости;
В — прирост сухого вещества плодов (y) томата в течение вегетационного периода (x);
1 — эмпирическая линия регрессии,
2 — логистическая кривая

В табл. 3.20 приведены следующие данные (по столбцам).

1. Пятидневки вегетационного периода (x). Интервалы между величинами аргумента могут быть и неравными.

2. Квадраты порядковых номеров пятидневок (x^2).

3. Прирост сухого вещества в плодах, в граммах на одно растение (y).

4. Расстояние между асимптотами a_1 (величина которого делится на y в столбце 4) определяется по графику эмпирических данных (рис. 24, В) так же, как и величина нижней асимптоты a_0 . В данном случае принимаем: $a_1 = 80,3$; $a_0 = 0$.

При равноотстоящих величинах аргумента значение верхней асимптоты можно определить также аналитически по формуле

$$a_1 = \frac{2b - y_1 - y_3}{b/y_2 - 1}, \quad (3.107)$$

где a_1 — расстояние между асимптотами; $b = (y_1 y_3) / y_2$; y_1, y_2, y_3 — три эмпирических значения функции, взятых через одинаковые интервалы аргумента, здесь $y_1 = 71,5$; $y_2 = 80,2$; $y_3 = 80,3$, подста-

Таблица 3.20. К вычислению коэффициентов логистического уравнения

		y	$\frac{80,3}{y}$	$\frac{80,3}{y} - 1$	$z = \lg \left(\frac{80,3}{y} - 1 \right)$		
1	2	3	4	5	6	7	8
(17)	(289)	80,3	1	0	—	—	—
(16)	(256)	80,3	1	0	—	—	—
15	225	80,2	1,0012	0,0012	3,0792	—2,9208	—43,812
14	196	77,6	1,03	0,03	2,4771	—1,5229	—21,3206
13	169	71,5	1,12	0,12	1,0792	—0,9208	—11,9704
12	144	67,8	1,18	0,18	1,2553	—0,7447	—8,9364
11	121	63,8	1,26	0,26	1,4150	—0,5850	—6,4350
10	100	58,7	1,37	0,37	1,5682	—0,4318	—4,3180
9	81	49,0	1,64	0,64	1,8062	—0,1938	—1,7442
8	64	41,7	1,93	0,93	1,9685	—0,0315	—0,2520
7	49	29,9	2,69	1,69	0,2279	0,2279	1,5953
6	36	26,5	3,03	2,03	0,3075	0,3075	1,8450
5	25	11,2	7,17	6,17	0,7903	0,7903	3,9515
4	16	5,28	15,2	14,2	1,1523	1,1523	4,6092
3	9	3,07	26,2	25,2	1,4014	1,4014	4,2042
2	4	1,20	66,9	65,9	1,8189	1,8189	3,6378
1	1	0,312	257	256	2,4082	2,4082	2,4082
120	1240	—	—	—	—	0,7552	—76,5374

вив которые в формулу (3.107), получим:

$$b = \frac{71,5 \cdot 80,3}{80,2} = 71,589; \quad a_1 = \frac{2 \cdot 71,589 - 71,5 - 80,3}{\frac{71,589}{80,2} - 1} = 80,3.$$

5. Из чисел столбца 4 вычитается единица.

6. Логарифмы десятичные (в искусственной форме) чисел столбца 5. Поскольку два верхних числа в этом столбце равны нулю, а логарифм нуля не существует, то приходится для сравнимости сумм в столбцах 1, 2, 7, 8 изъять также два верхних числа в столбцах 1 и 2 (они взяты в скобки), которые при суммировании не учитываются, причем число наблюдений соответственно также сокращается до $N = 15$.

7. Перевод логарифмов из столбца 6 в естественную форму для удобства арифметических действий с ними.

8. Произведения чисел 1 и 7 столбцов.

После этого суммируем числа в столбцах 1, 2, 7, 8, находим средние:

$$M_x = \frac{\sum x}{N} = \frac{120}{15} = 8,0; \quad M_z = \frac{\sum z}{N} = \frac{0,7552}{15} = 0,05035$$

и коэффициенты уравнения по формулам:

$$\beta = \frac{\frac{\sum xz}{M_x} - \sum z}{\frac{\sum x^2}{M_x} - \sum x}, \quad (3.108)$$

$$\gamma = M_z - \beta M_x, \quad (3.109)$$

по которым:

$$\beta = \frac{\frac{-76,5374}{8} - 0,7552}{\frac{1240}{8} - 120} = -0,2949;$$

$$\gamma = 0,05035 + 0,2949 \cdot 8 = 2,4096.$$

Для проверки коэффициенты уравнения можно найти также, решив систему:

$$\begin{aligned} N\gamma + \beta \sum x &= \sum z, \\ \gamma \sum x + \beta \sum x^2 &= \sum xz, \end{aligned} \quad (3.110)$$

для чего подставляем в нее известные числа: $15\gamma + \beta 120 = 0,7552$, $\gamma 120 + \beta 1240 = -76,5374$, откуда получаем те же значения коэффициентов.

Уравнение (3.106) приобретает для рассматриваемого примера следующую конкретную форму:

$$y' = \frac{80,3}{1 + 10^{2,4096 - 0,2949x}} + 0,$$

где y' — теоретические значения функции, которые вычислены в табл. 3.21; в ней по столбцам приведены следующие данные.

1. Аргумент — пятидневки вегетационного периода (x).
2. Значение коэффициента $\beta = -0,2949$ умножено на значения аргумента x .
3. К величинам из предыдущего столбца прибавлено значение коэффициента $\gamma = 2,4096$, результат есть логарифм в естественной форме.
4. Логарифмы чисел столбца 3 в искусственной форме.
5. Антилогарифмы чисел столбца 4.
6. Числа столбца 5, увеличенные на 1.
7. Расстояние между асимптотами $a_1 = 80,3$, деленное на числа из столбца 6, что в данном примере соответствует теоретическим значениям функции (y').
8. Эмпирические значения функции (y).
9. Разница соответствующих эмпирических и теоретических значений функции. Чем меньше сумма чисел этого столбца, тем лучше качество аппроксимации.

Таблица 3.21. Вычисление значений функции по уравнению логистической кривой

	βx	$\gamma + \beta x$	$\lg (10^{\gamma + \beta x})$	$10^{\gamma + \beta x}$	$1 + 10^{\gamma + \beta x}$	y'	y	$y - y'$	$(y - y')^2$
1	2	3	4	5	6	7	8	9	10
17	-5,0133	-2,6037	3,3963	0,00249	1,00249	80,1	80,3	0,2	0,04
16	-4,7184	-2,3088	3,6912	0,004911	1,004911	79,9	80,3	0,4	0,16
15	-4,4235	-2,0139	3,9861	0,009684	1,009684	79,5	80,2	0,7	0,49
14	-4,1286	-1,7190	2,2810	0,0191	1,0191	78,8	77,6	-1,2	1,44
13	-3,8337	-1,4241	2,5759	0,03766	1,03766	77,4	71,5	-5,9	34,81
12	-3,5388	-1,1292	2,8708	0,07426	1,07426	74,7	67,8	-6,9	47,61
11	-3,2439	-0,8343	1,1657	0,1464	1,1464	70,0	63,8	-6,2	38,44
10	-2,9490	-0,5394	1,4606	0,2887	1,2887	62,3	58,7	-3,6	12,96
9	-2,6541	-0,2445	1,7555	0,5695	1,5695	51,2	49,0	-2,2	4,84
8	-2,3592	0,0504	0,0504	1,123	2,123	37,8	41,7	3,9	15,21
7	-2,0643	0,3453	0,3453	2,215	3,215	25,0	29,9	4,9	24,01
6	-1,7694	0,6402	0,6402	4,367	5,367	15,0	26,5	11,5	132,25
5	-1,4745	0,9351	0,9351	8,611	9,611	8,35	11,2	2,85	8,12
4	-1,1796	1,2300	1,2300	17,02	18,02	4,46	5,28	0,82	0,67
3	-0,8847	1,5249	1,5249	33,49	34,49	2,33	3,07	0,74	0,55
2	-0,5898	1,8198	1,8198	66,04	67,04	1,20	1,20	0	0
1	-0,2949	2,1147	2,1147	130,2	131,2	0,612	0,312	-0,3	0,09
								-0,29	321,69

10. Квадраты чисел из столбца 9, сумма которых $\sum (y - y')^2 = 321,69$.

Приведенный алгоритм вычисления теоретических значений функции кратко может быть записан в виде логарифмированного уравнения (3.106):

$$\lg\left(\frac{a_1}{y' - a_0} - 1\right) = \gamma + \beta x,$$

где в нашем примере $a_0 = 0$.

Вычисление y' начинается с нахождения значения выражения $\gamma + \beta x$, которое представляет собой логарифм в естественной форме, далее переводим этот логарифм в искусственную форму и находим по нему число, к которому прибавляем единицу. На полученное выражение делим a_1 , что и дает значение y'

Например, y' для $x = 17$, по уравнению:

$$\lg\left(\frac{80,3}{y'} - 1\right) = 2,4096 - 0,2949x;$$

$$2,4096 - 0,2949 \cdot 17 = -2,6037 = \bar{3},3963,$$

что соответствует числу 0,00249, к которому прибавляем 1 и делим:

$$80,3/1,00249 = 80,1 = y'$$

Ошибка найденного уравнения аппроксимации прироста сухого вещества в плодах томата в зависимости от номера пятидневки его вегетационного периода вычисляется по формуле (3.04):

$$m_{y \cdot x} = \sqrt{\frac{321,69}{17 - 3}} = 4,7936 \text{ г.}$$

На рис. 24, в изображены эмпирическая и теоретическая линии регрессии прироста сухого вещества в плодах томата. При анализе логистического уравнения важную роль играют точки перегиба кривой и ее пересечения с осями.

Точка пересечения логистической кривой с осью ординат равна:

$$y_{\Pi} = \frac{a_1}{1 + 10^{\gamma}} + a_0, \quad (3.111)$$

координаты средней точки перегиба кривой:

$$x_{\Sigma} = -\frac{\gamma}{\beta}, \quad y_{\Sigma} = \frac{a_1}{2} + a_0, \quad (3.112)$$

где обозначения те же, что и к формуле (3.106).

Для рассматриваемого примера точка пересечения кривой с осью ординат:

$$y_{\Pi} = \frac{80,3}{1 + 10^{2,4096}} + 0 = 0,612;$$

координаты точки перегиба кривой:

$$x_{\Sigma} = -\frac{2,4096}{-0,2949} = 8,17; \quad y_{\Sigma} = \frac{80,3}{2} + 0 = 40,15.$$

При более детальном анализе полезно определить координаты также нижней и верхней точек перегиба логистической кривой:

$$x_H = x_C - \frac{0,5719}{|\beta|}; \quad y_H = 0,2113a_1; \quad (3.113)$$

$$x_B = x_C + \frac{0,5719}{|\beta|}; \quad y_B = 0,7887a_1; \quad (3.114)$$

где x_H, y_H — абсцисса и ордината нижней точки перегиба; x_B, y_B — абсцисса и ордината верхней точки перегиба; x_C — абсцисса средней точки перегиба; a_1 — расстояние между асимптотами; $|\beta|$ — коэффициент (при x) уравнения без знака.

Для нашего примера:

$$x_H = 8,17 - 0,5719/0,2949 = 6,23; \quad y_H = 0,2113 \cdot 80,3 = 16,97.$$

$$x_B = 8,17 + 0,5719/0,2949 = 10,11; \quad y_B = 0,7887 \cdot 80,3 = 63,33.$$

Для проверки вычислений ординат точек перегиба можно применить равенство:

$$a_1 = y_H + y_B, \quad 80,3 = 16,97 + 63,33. \quad (3.115)$$

Для проверки вычислений абсцисс верхней и нижней точек перегиба служит равенство:

$$2x_C = x_H + x_B, \quad 2 \cdot 8,17 = 6,23 + 10,11. \quad (3.116)$$

Таким образом, в начале седьмой (6,23), девятой (8,17) и одиннадцатой (10,11) пятидневок в процессе накопления сухих веществ в плодах томата происходят качественные изменения, которые подлежат биологическому объяснению. По-видимому, эти явления связаны с внешними и внутренними признаками созревания, а также с изменениями в биохимическом составе веществ плода томата. Кроме того, эти данные при аналогичных метеоусловиях и агротехнике позволяют прогнозировать урожай помидор по сухому веществу. Для этого, на 31-й день после начала роста плодов следует определить в них содержание сухого вещества (y_H) и разделить его на 0,21, что даст величину содержания сухого вещества в конце вегетационного периода (a_1), которую нетрудно пересчитать при желании на вес сырой массы. Например: $a_1 = y_H/0,21 = 16,97/0,21 = 80,3$. Этот прогноз можно повторить и уточнить его на 41-й день от начала роста плодов по формуле: $a_1 = y_C/0,5$, и на 51-й день по формуле: $a_1 = y_B/0,79$. Значение точки пересечения кривой с осью ординат (y_H) в данном примере заключается в том, что существовал минимум сухой массы (семена, рассада) перед тем, как началось ее накопление в процессе роста. В этом же состоит и теоретическое значение нижней асимптоты. Значение верхней асимптоты в нашем примере указывает предел, выше которого при существующих экологических условиях, ассортименте сортов и агротехнике невозможно получение содержания сухого вещества в плодах томата.

В некоторых случаях качество аппроксимации логистической кривой может повысить применение в качестве показателя степени в ее уравнении не прямой линии, а параболы.

Таблица 3.22. Вычисление коэффициентов уравнения логистической кривой

Годы, x	Всхо- жесть %, y		x_0^2	x_0^4	$z_0 x_0^2$		y'
1	2	3	4	5	6	7	8
7	0,1	6,7593	14,6092	213,4287	98,4675	25,762	0,015
6	0,64	4,8968	7,9648	63,438	38,8491	13,7656	0,439
5	6	2,594	3,3204	11,0251	8,5494	4,6918	6,98
4	36	0,33446	0,676	0,457	0,2131	0,2592	40,86
3	52	-0,41612	0,0316	0,001	-0,0138	0,0774	72,95
2	71,4	-1,5667	1,3872	1,9243	-2,2	1,8679	82,14
1	80,3	-2,5943	4,7428	22,4941	-12,3953	5,6917	84,31
0,5	85	-4,1799	7,1706	51,4175	-30,1101	11,2443	84,64
0,1	86	-5,655	9,4729	89,7358	-53,7511	17,4641	84,73
28,6		0,17254	49,3755	453,9215	47,6088	80,824	

Рассмотрим пример расчета коэффициентов уравнения такой кривой вида:

$$y' = \frac{a_1}{1 + e^{\gamma + \beta x + \alpha x^2}} + a_0, \quad (3.117)$$

где y' — теоретические значения функции; a_1 , a_0 — верхняя и нижняя асимптоты; α , β , γ — коэффициенты. Требуется аппроксимировать данные о всхожести семян жимолости через разные сроки их хранения. В табл. 3.22 по столбцам приведены следующие данные.

1. Сроки хранения семян, в годах.

2. Всхожесть семян, в процентах.

3. Величину верхней асимптоты делим поочередно на эмпирические значения функции, т. е. всхожести без величины нижней асимптоты, вычитаем единицу из частного и берем натуральный логарифм полученного числа, например

$$z = \ln \left(\frac{a_1}{y - a_0} - 1 \right) = \ln \left(\frac{86,3}{0,1 - 0} - 1 \right) = 6,7593.$$

Значение верхней асимптоты принимаем $a_1 = 86,3$, т. е. немного больше максимального y , а величина a_0 здесь равна нулю, что соответствует естественному значению всхожести при полной потере жизнеспособности семян в результате слишком длительного их хранения.

4. Квадраты преобразованных значений аргумента, из которых вычитается его средняя арифметическая, т. е. $x_0 = x - M_x$, где, как обычно, $M_x = \Sigma x / N = 28,6 / 9 = 3,1778$, например, $x_0 = 7 - 3,1778 = 3,8222$; последнее число в квадрате равно 14,6092.

5. Преобразованные значения аргумента в четвертой степени, т. е. $x_0^4 = (x_0^2)^2 = (14,6092)^2 = 213,4287$.

6. Величины из столбца 3 $z = \ln \left(\frac{a_1}{y - a_0} - 1 \right)$ преобразуем по тому же принципу, что и в столбце 4, т. е. находим среднюю ариф-

метическую: $M_z = \Sigma z/N = 0,17254/9 = 0,01917$ и вычитаем ее из всех значений z , например: $z_0 = z - M_z = 6,7593 - 0,01917 = 6,740$. Полученные z_0 умножаем попарно на величины x_0^2 из столбца 4.

7. Произведения преобразованных величин функции и аргумента, например $z_0 = 6,740$ и $x_0 = 3,8222$; $z_0 x_0 = 6,740 \cdot 3,8222 = 25,762$. Далее находим суммы столбцов 3—7 и коэффициенты уравнения логистической кривой по формулам:

$$\alpha = \frac{\sum z_0 x_0^2}{\sum x_0^4 - \frac{(\sum x_0^2)^2}{N}}, \quad (3.118)$$

$$\beta = \frac{\sum x_0 z_0}{\sum x_0^2} - 2\alpha M_x, \quad (3.119)$$

$$\gamma = M_z - \frac{\alpha \sum x_0^2}{N} - M_x (\beta + \alpha M_x), \quad (3.120)$$

$$\alpha = \frac{47,6088}{453,9215 - 2437,9499/9} = 0,2601;$$

$$\beta = \frac{80,824}{49,3755} - 2 \cdot 0,2601 \cdot 3,1778 = -0,0162;$$

$$\gamma = 0,0192 - \frac{0,2601 \cdot 49,3755}{9} - 3,1778(-0,0162 + 0,2601 \cdot 3,1778) = -3,9906.$$

8. Вычисление теоретических значений функции начинаем с выражения:

$$\gamma + x (\beta + \alpha x) = -3,9906 + x (-0,0162 + 0,2601 x),$$

затем берем экспоненту от полученного числа, прибавляем к нему единицу, делим на него величину верхней асимптоты и прибавляем величину нижней асимптоты, т. е. вычисляем по уравнению (3.117), которое получило следующие коэффициенты:

$$y' = \frac{86,3}{1 + e^{-3,9906 + x(-0,0162 + 0,2601x)}} + 0.$$

Ошибку полученного уравнения находим по формуле (3.04)

$$m_{y \cdot x} = \sqrt{\frac{596,7}{9-3}} = 10\%.$$

График рассматриваемой зависимости приведен в § 5.05, где достигнута более близкая аппроксимация иным способом. Значения аргумента по значениям функции можно вычислить преобразуя

уравнение (3.117) следующим образом¹:

$$x = \frac{-\beta \pm \sqrt{\beta^2 + 4\alpha \left(\ln \left(\frac{a_1}{y - a_0} - 1 \right) - \gamma \right)}}{2\alpha}. \quad (3.121)$$

Критические точки кривой:

$$y = \frac{a_1}{1 + e^{\gamma + \beta x + \alpha x^2}} + a_0,$$

находим следующим образом. Координаты верхней точки максимума равны:

$$x_M = -\frac{\beta}{2\alpha}, \quad (3.122)$$

$$y_M = \frac{a_1}{1 + e^{\gamma + \beta x_M + \alpha x_M^2}} + a_0, \quad (3.123)$$

или:

$$y_M = \frac{a_1}{1 + e^{\gamma - \frac{\beta^2}{4\alpha}}} + a_0.$$

Вторая производная уравнения (3.117) равна:

$$y'' = a_1 z \left(\frac{2\alpha x + \beta}{1 + z} \right)^2 \left[2 \left(\frac{z}{1 + z} - \frac{\alpha}{(2\alpha x - \beta)^2} \right) - 1 \right], \quad (3.124)$$

где $z = e^{\gamma + x(\beta + \alpha x)}$. Значение x , при котором y'' принимает максимальное значение, соответствует абсциссе критической точки Н, значение x , соответствующее $y'' = 0$, равно абсциссе точки С и значение x , при котором y'' принимает минимальное значение, соответствует критической точке В. Ординаты точек В, Н находим подобно тому, как и ординату y_M , по уравнению (3.123). При отрицательных и положительных значениях x могут быть по две симметрично расположенных точки Н, С, В, так как кривая становится в этом случае двускатной (рис. 25). Придавая аргументу ряд последовательных значений с небольшим шагом, выявляем абсциссы критических точек, которые в данном примере равны: $x_B = 3,4584$; $x_C = 4,06314$; $x_H = 4,6809$; затем по этим абсциссам находим ординаты: $y_B = 61,98$; $y_C = 38,05$; $y_H = 14,11$. В рассматриваемом примере отрицательные величины x не имеют смысла, поэтому не определяем координат критических точек из левой части кривой на рис. 25. В тех случаях, когда вычисление их желательно, для проверки можно использовать равенство:

$$\left. \begin{aligned} x_{B-} + x_{B+} \\ x_{C-} + x_{C+} \\ x_{H-} + x_{H+} \end{aligned} \right\} = \beta/\alpha, \quad (3.125)$$

¹ При дальнейшем анализе логистической кривой автор пользовался консультацией сотрудника ВЦ АН СССР С. В. Сорокина, которому он выражает свою благодарность.

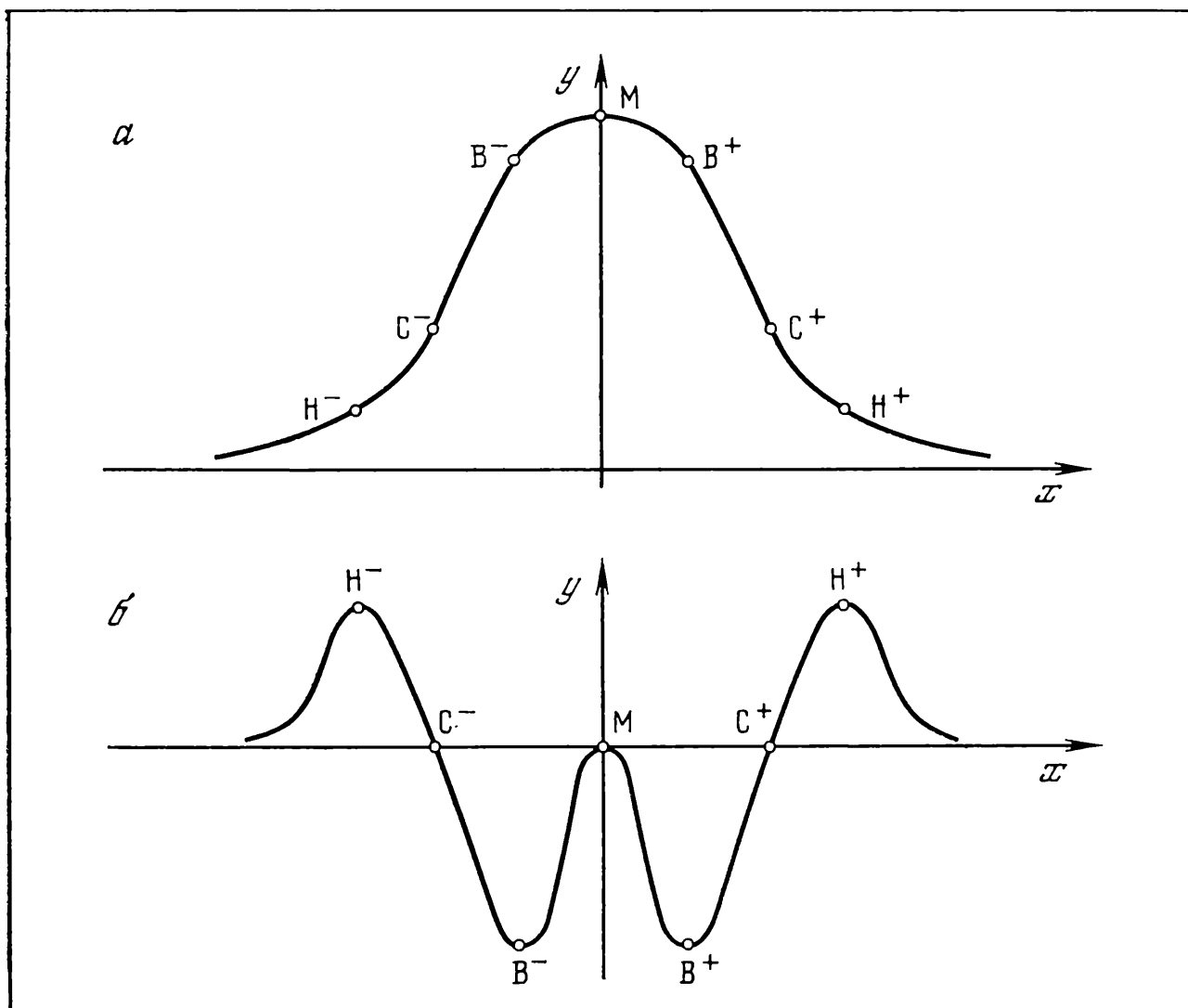


Рис. 25. Схематический график кривой (3.117) (а) и график ее второй производной (3.124) (б)

М, В, С, Н — критические точки

Координаты точки максимума M находим по формулам (3.122), (3.123):

$$x_M = -\frac{-0,0162}{2 \cdot 0,2601} = 0,03115; \quad y_M = \frac{86,3}{1 + e^{\frac{-3,9906 - \frac{(-0,0162)^2}{4 \cdot 0,2601}}}} = 84,73.$$

Биологическое значение найденных координат поясняется в § 5.05.

§ 3.12. Кривая Гаусса

Для аппроксимации таких эмпирических зависимостей, точки которых располагаются по колоколообразной кривой и по краям плавно приближаются к оси абсцисс, можно использовать уравнение кривой Гаусса:

$$y = be^{-h^2(x-a)^2}, \quad (3.126)$$

где y — зависимая переменная; b , h , a — коэффициенты; x — независимая переменная; $e = 2,71828$.

Логарифмируем (3.126) и составим систему уравнений по трем избранным точкам:

$$\ln y_1 = \ln b - h^2 (x_1^2 - 2ax_1 + a^2),$$

$$\ln y_2 = \ln b - h^2 (x_2^2 - 2ax_2 + a^2),$$

$$\ln y_3 = \ln b - h^2 (x_3^2 - 2ax_3 + a^2).$$

Решая эту систему относительно искомых коэффициентов, получаем:

$$h^2 = \frac{\frac{\ln(y_3/y_2)}{x_2 - x_3} - \frac{\ln(y_2/y_1)}{x_1 - x_2}}{x_3 - x_1}, \quad (3.127)$$

$$a = 0,5 \left[x_1 + x_2 - \frac{\ln(y_2/y_1)}{h^2(x_1 - x_2)} \right], \quad (3.128)$$

$$b = y_1 e^{h^2 x_1^2 - 2ax_1 + a^2} \quad (3.129)$$

Пример. Найдем коэффициенты уравнения (3.126) применительно к эмпирической зависимости, приведенной в табл. 3.23. При избранных точках: $x_1 = 37$, $x_2 = 79$, $x_3 = 121$, $y_1 = 2$, $y_2 = 171$, $y_3 = 16$, коэффициенты по формулам (3.127)–(3.129) равны:

$$h^2 = 0,0019325; \quad a = 85,405; \quad b = 185,10.$$

Теоретические точки кривой по уравнению: $y' = 185,10 e^{-0,0019325(x-85,405)^2}$ приведены в столбце 4 табл. 3.23. Сравнивая значения зависимой переменной, полученные по уравнению, с исходными данными (столбец 2) и со значениями нормальной кривой, полученными точным способом (столбец 3), видим, что их совпадение недостаточное. В нижних строках таблицы приводятся величины хи-квадрата (χ^2) и его оценка по коэффициенту Романовского (КР) (см. § 4.14) для каждого теоретического ряда ($y' - y'''$),

Таблица 3.23. Данные к вычислению коэффициентов уравнения кривой Гаусса

	y	f_H	y'	y'	y'''
1	2	3	4	5	6
37	2	4,5	2	3,6	6,9
51	29	31,6	18,8	29,0	40,8
65	119	106,1	82,8	104,8	119,0
79	171	170,4	171,0	171,0	171,0
93	121	130,6	165,5	126,1	121,0
107	42	47,8	75,2	42,0	42,2
121	16	8,3	16,0	6,3	7,2
χ^2			105,44	17,78	17,65
КР:			35,86	4,87	4,83

полученного по избранным точкам. Аппроксимацию можно улучшить, взяв для подстановки в формулы другие три избранные точки: $x_1 = 51$, $x_2 = 79$, $x_3 = 107$, $y_1 = 29$, $y_2 = 171$, $y_3 = 42$, находящиеся ближе к вершине кривой, в соответствии с которыми получим уравнение: $y'' = 171,91 e^{-0,0020271(x-80,630)^2}$.

Теоретические точки кривой, вычисленные по этому уравнению, приведены в столбце 5 табл. 3.23, они удовлетворительно совпадают с эмпирическими данными и с теоретическими точными значениями (столбец 3). Совпадение еще более улучшается, если взять избранные точки на вершине кривой; в столбце 6 приведены значения функции, вычисленные по избранным точкам: $x_1 = 65$, $x_2 = 79$, $x_3 = 93$ и по уравнению: $y''' = 171,02 e^{-0,0018072(x-79,165)^2}$.

Для проверки этого предположения были вычислены теоретические ряды по трем крайним точкам с одной ($x_1 = 93$, $x_2 = 107$, $x_3 = 121$) и с другой стороны ($x_1 = 37$, $x_2 = 51$, $x_3 = 65$) кривой. Совпадение в обоих случаях оказалось наихудшим ($\chi^2 = 3001,9$; $KP = 1059,9$ и близко к этому во втором случае).

Таким образом, лучшее совпадение теоретических рядов с эмпирическими достигается, если брать избранные точки возможно ближе к вершине кривой. Это и понятно, имея в виду наибольшую достоверность частот нормальной кривой в ее середине, в области математического ожидания.

Методом избранных точек можно также вычислить среднюю арифметическую и дисперсию эмпирического распределения, близкого к нормальному, для чего исходное уравнение (см. § 1.25):

$$f = \frac{Nc}{\sigma \sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}}$$

логарифмируем и напишем систему уравнений для четырех точек:

$$\ln f_1 = \ln N + \ln c - \ln \sigma - \ln m - d (x_1 - M)^2,$$

$$\ln f_2 = \ln N + \ln c - \ln \sigma - \ln m - d (x_2 - M)^2,$$

$$\ln f_3 = \ln N + \ln c - \ln \sigma - \ln m - d (x_3 - M)^2,$$

$$\ln f_4 = \ln N + \ln c - \ln \sigma - \ln m - d (x_4 - M)^2,$$

где $m = \sqrt{2\pi}$; f — частоты.

Обозначим: $d = 1/(2\sigma^2)$, $a = f_1/f_2$, $b = f_3/f_4$.

Решая систему четырех уравнений относительно коэффициентов, получим формулы:

$$M = \frac{(x_2^2 - x_1^2) \ln b - (x_4^2 - x_3^2) \ln a}{2 [(x_2 - x_1) \ln b - (x_4 - x_3) \ln a]}, \quad (3.130)$$

$$d = \frac{\ln a}{(x_2 - M)^2 - (x_1 - M)^2}, \quad (3.131)$$

$$\sigma = 1/2d. \quad (3.132)$$

Приняв (из столбцов 1 и 2 табл. 3.23) за избранные точки: $x_1 = 37$, $x_2 = 65$, $x_3 = 93$, $x_4 = 121$, $f_1 = 2$, $f_2 = 119$, $f_3 = 121$, $f_4 = 16$, по формулам (3.130)–(3.132) получим: $M = 88,4$; $\sigma^2 = 256,7$; $\sigma = 16,0$. Точные значения этих параметров для этого же ряда: $M = 80,960$; $\sigma^2 = 265,06$; $\sigma = 16,281$.

Вершина нормальной кривой находится в точке с координатами: $x = M$; $y = f_{\max} = \frac{Nc}{\sigma\sqrt{2\pi}}$; абсцисса точки перегиба кривой равна σ .

Формулы (3.130)–(3.132) можно применять в тех случаях, когда вариационный ряд по каким-либо причинам оказался неполным; для применения изложенного метода достаточно, чтобы в ряду сохранилось четыре варианта с их частотами.

§ 3.13. Логарифмическая функция

Кривые этого типа графически отражают такие зависимости, в которых корреляция двух факторов, довольно сильная в начале, становится слабой и почти незначимой в конце данного процесса. Технику вычисления коэффициентов уравнения логарифмической кривой:

$$y = a + bx + d \ln x \quad (3.133)$$

рассмотрим на примере аппроксимации эмпирической зависимости срока созревания семян от начала цветения у видов травянистых многолетников.

1. В первых двух столбцах табл. 3.24 даны сроки этих двух фенофаз в днях от 1 марта. Предстоящие вычисления с этими главным образом трехзначными числами могли бы стать затруднительными, если бы не были преобразованы исходные величины с целью снижения их порядка (столбцы 3 и 4). Заметим, что подобное преобразование желательно делать при аппроксимации любыми типами

Таблица 3.24. Сроки зацветания (x) и созревания с мян (y) у видов травянистых многолетников в Москве

В днях от 1 марта		Кодированные даты		Теоретические значения y'
Зацветание, x'	Созревание семян, y''	$x = \frac{x' - 20}{20}$	$y = \frac{y''}{100}$	
40	97,5	1	0,975	0,83283
60	124,4	2	1,244	1,2625
80	138,1	3	1,381	1,5141
100	152,8	4	1,528	1,6926
120	174,2	5	1,742	1,8313
140	198,9	6	1,989	1,9448
160	212,05	7	2,1205	2,0408
180	220,2	8	2,202	2,1239
200	225,8	9	2,258	2,1973

$N = 9$; $A = 20$; $c = 20$ 45

кривых, особенно тогда, когда исходные величины при расчетах возводят в степень или логарифмируют.

Для приведения ряда аргумента к натуральному ряду чисел следует найти подходящие значения величин A (начального значения x) и c (интервала между значениями x). В затруднительных случаях можно с этой целью решить систему уравнений:

$$x_1 = \frac{x'_1 - A}{c}, \quad x_2 = \frac{x'_2 - A}{c},$$

например, для данных в столбце 1 табл. 3.24:

$$1 = (40 - A)/c; \quad 2 = (60 - A)/c; \quad \text{откуда } A = 20; \quad c = 20.$$

Эта несложная операция существенно снижает трудоемкость расчетов и повышает их точность.

1. Найдем суммы величин: x , y , $x \ln x$, $\ln x$, xy , $y \ln x$, x^2 , $\ln^2 x$, для чего выполним соответствующие вычисления с числами, приведенными в столбцах 3, 4 табл. 3.24.

2. Средние арифметические обоих рядов равны:

$$M_x = (\Sigma x)/N = 45/9 = 5; \quad M_y = (\Sigma y)/N = 15,44/9 = 1,7155.$$

3. Вспомогательные величины определим по следующим формулам:

$$A = \frac{\sum (x \ln x)}{\sum \ln x} = \frac{79,057}{12,802} = 6,1754;$$

$$B = \frac{\sum xy}{\sum x} = \frac{87,144}{45} = 1,9365;$$

$$B = \frac{\sum y \ln x}{\sum \ln x} = \frac{24,532}{12,802} = 1,9163;$$

$$\Gamma = \frac{\sum x^2}{\sum x} = \frac{285}{45} = 31,666; \tag{3.134}$$

$$Д = \frac{\sum x \ln x}{\sum x} = \frac{79,057}{45} = 1,7568;$$

$$E = \frac{\sum \ln x}{N} = \frac{12,802}{9} = 1,4224;$$

$$Ж = \frac{\sum \ln^2 x}{\sum \ln x} = \frac{22,348}{12,802} = 1,7457.$$

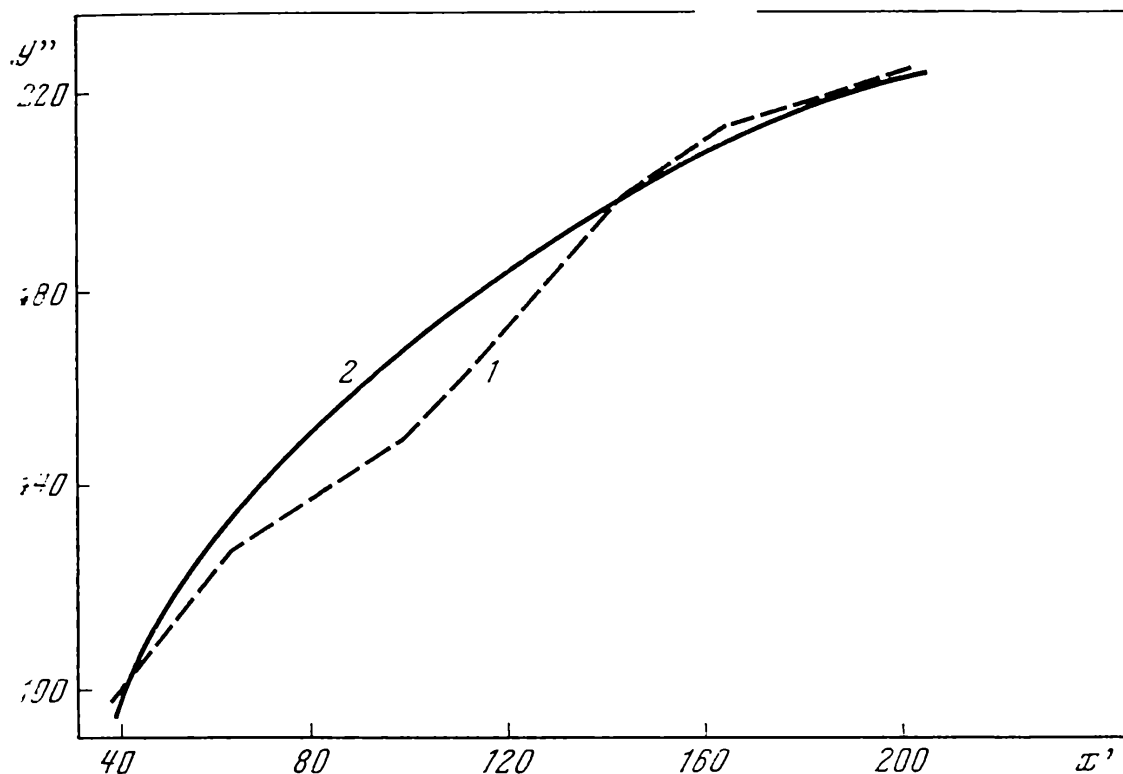


Рис. 26. Аппроксимация зависимости сроков созревания семян (y'') у видов травянистых многолетников от даты начала их цветения (x')

1 — эмпирические данные; 2 — логарифмическая кривая

Рис. 27. Формы кривых первой группы сложности

а) $y = ae^{bx}$ (см. § 3.09)

- 1) $y = e^{-0,69x}$
- 2) $y = e^{-0,51x}$
- 3) $y = e^{-0,36x}$
- 4) $y = e^{-0,22x}$
- 5) $y = e^{-0,105x}$
- 6) $y = e^{-0,05x}$
- 7) $y = e^{-0,01x}$
- 8) $y = e^{0,01x}$
- 9) $y = e^{0,05x}$
- 10) $y = e^{0,09x}$
- 11) $y = e^{0,18x}$

б) $y = 1/(a + bx)$

- 1) $y = 1/(x + 1)$
- 2) $y = 1/(x - 1)$
- 3) $y = 1/(0,5x - 1)$
- 4) $y = 1/(7 - x)$
- 5) $y = 1/(8 - 0,6x)$
- 6) $y = 1/(10 - 0,5x)$

в) $y = x/(a + bx)$

- 1) $y = x/(0,2 + 0,2x)$
- 2) $y = 1/(0,1 + 0,2x)$
- 3) $y = x/(-0,1 + 0,2x)$
- 4) $y = x/(-0,2 + 0,2x)$

г) $y = ax^b + c$

- 1) $y = 1 + 0,5x^{,5}$
- 2) $y = 1 - 0,2x^{0,5}$
- 3) $y = 1 + 0,04x^2$
- 4) $y = 1 + 0,08x^2$
- 5) $y = 1 + 2x^{-0,5}$
- 6) $y = 1 + 3x^{-0,5}$

д) $y = ae^{bx} + c$

- 1) $y = 2 - e^{-0,6x}$
- 2) $y = 2 - e^{-0,51x}$
- 3) $y = 2 - e^{-0,36x}$
- 4) $y = 2 - e^{-0,22x}$
- 5) $y = 2 - e^{-0,16x}$

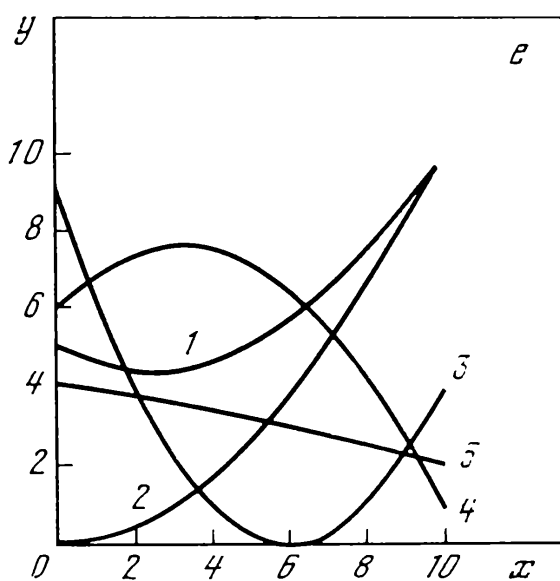
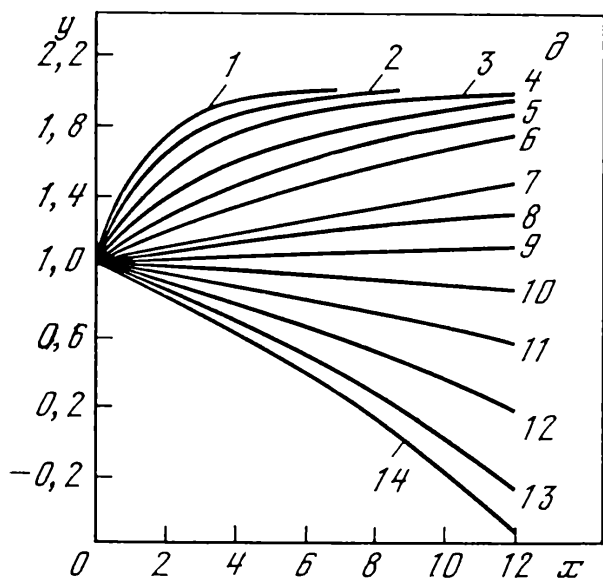
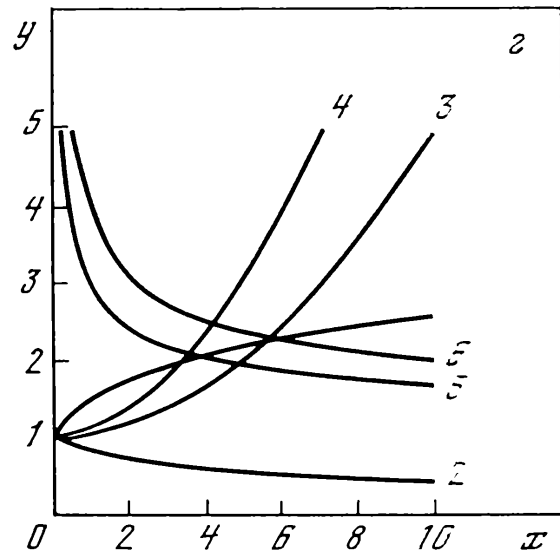
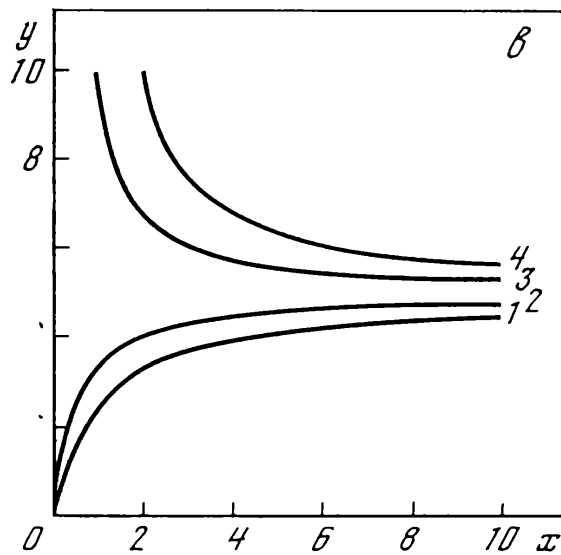
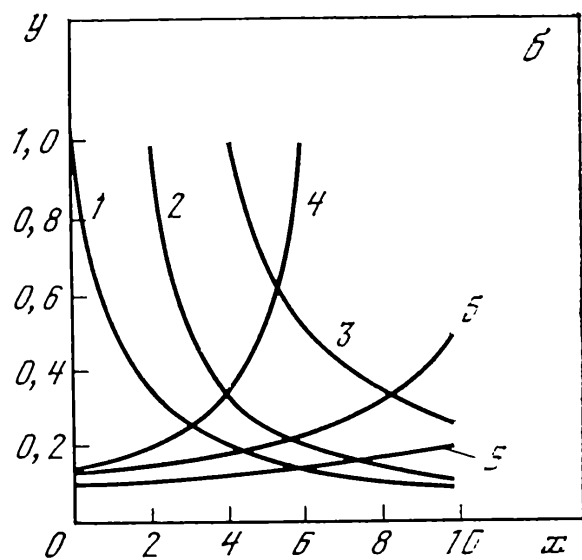
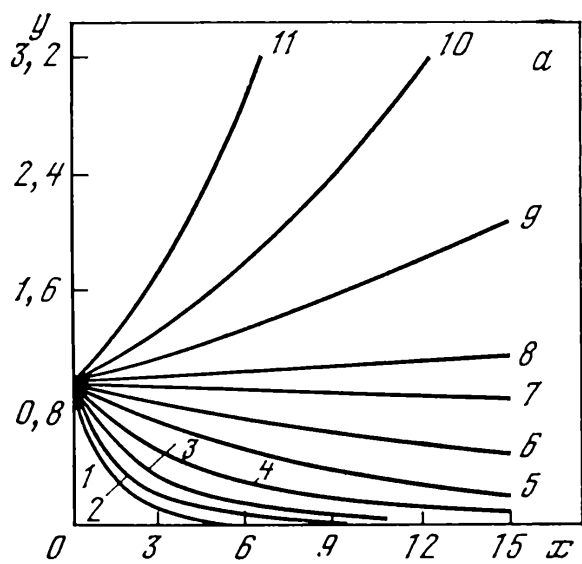
д) $y = ae^{bx} + c$

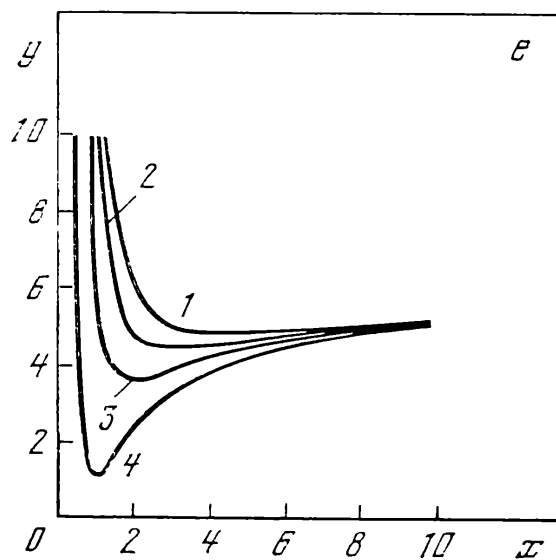
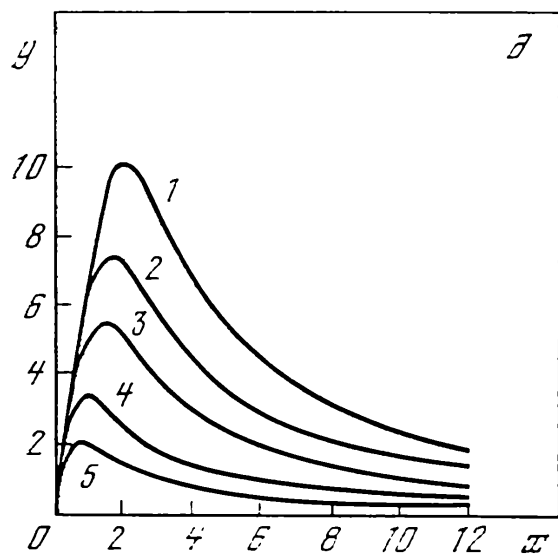
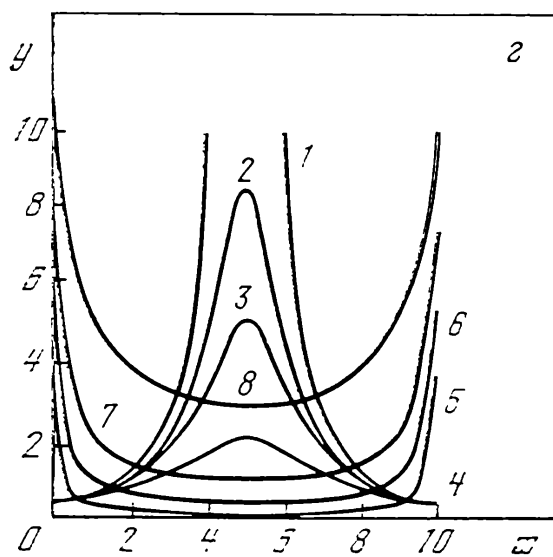
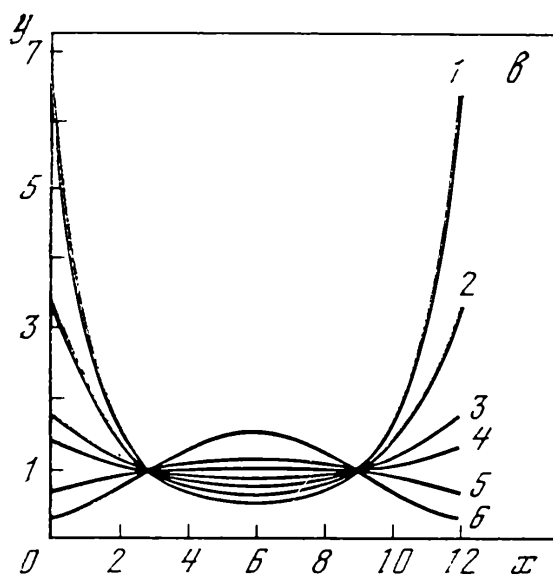
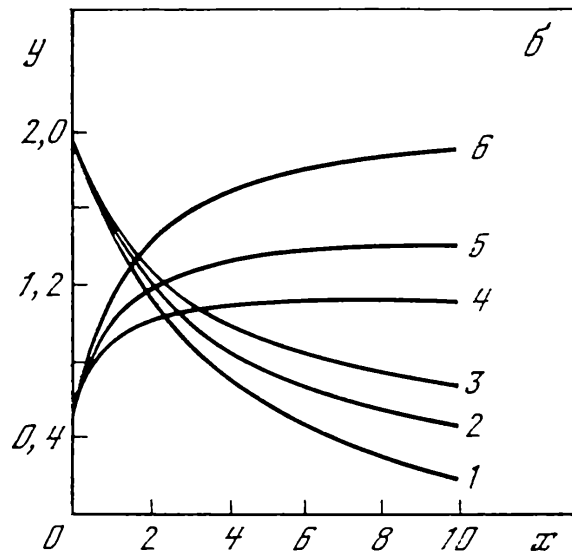
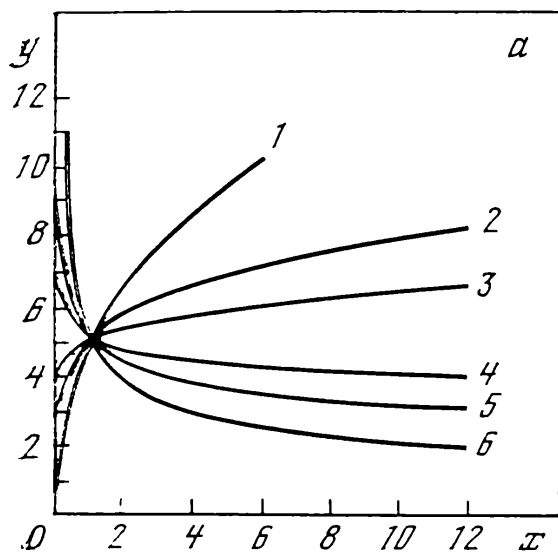
- 6) $y = 2 - e^{-0,105x}$
- 7) $y = 2 - e^{-0,05x}$
- 8) $y = 2 - e^{-0,03x}$
- 9) $y = 2 - e^{-0,01x}$
- 10) $y = 2 - e^{0,01x}$
- 11) $y = 2 - e^{0,03x}$
- 12) $y = 2 - e^{0,05x}$
- 13) $y = 2 - e^{0,068x}$
- 14) $y = 2 - e^{0,077x}$

е) $y = a + bx + cx^2$

(см. § 3.03, 3.06)

- 1) $y = 5 - 0,5x + 0,1x^2$
- 2) $y = 0,1x^2$
- 3) $y = 9 - 3x + 0,25x^2$
- 4) $y = 6 + x - 0,15x^2$
- 5) $y = 4 - 0,1x - 0,01x^2$





4. Коэффициенты уравнения логарифмической кривой находим по формулам:

$$d = \frac{A(B - M_y) + M_x(B - B) + \Gamma(M_y - B)}{A(D - E) + M_x(K - D) + \Gamma(E - K)} =$$

$$= \frac{6,1754(1,9365 - 1,7155) + 5(1,9163 - 1,9365) + 31,666(1,7155 - 1,9163)}{6,1754(1,7568 - 1,4224) + 5(1,7457 - 1,7568) + 31,666(1,4224 - 1,7457)} =$$

$$= 0,61911; \quad (3.135)$$

$$b = \frac{d(D - E) + M_y - B}{M_x - \Gamma} =$$

$$= \frac{0,61911(1,7568 - 1,4224) + 1,7155 - 1,9365}{5 - 31,666} = 0,00052501; \quad (3.136)$$

$$a = M_y - bM_x - dE = 1,7155 - 0,00052501 \cdot 5 - 0,61911 \cdot 1,4224 = 0,8323. \quad (3.137)$$

5. Вычисленные по уравнению:

$$y' = 0,8323 + 0,00052501x + 0,61911 \ln x,$$

Рис. 28. Формы кривых второй группы сложности

а) $y = ax^b$ (см. § 3.07)

- 1) $y = 2x^{-2}$
- 2) $y = 2x^{-1}$
- 3) $y = 2x^{-0,5}$
- 4) $y = 2x^{0,24}$
- 5) $y = 2x^{0,5}$
- 6) $y = 2x$
- 7) $y = 2x^{1,5}$
- 8) $y = 2x^2$

б) $y = x/(a + bx) + c$

- 1) $y = 2 - x/(1,5 + 0,4x)$
- 2) $y = 2 - x/(1,5 + 0,5x)$
- 3) $y = 2 - x/(1,5 + 0,6x)$
- 4) $y = 0,5 + x/(1 + 1,5x)$
- 5) $y = 0,5 + x/(1 + x)$
- 6) $y = 0,5 + x/(1 + 0,6x)$

в) $\lg y = a + bx + cx^2$

- 1) $\lg y = 0,81 - 0,36x + 0,03x^2$
- 2) $\lg y = 0,54 - 0,24x + 0,02x^2$
- 3) $\lg y = 0,27 - 0,12x + 0,01x^2$
- 4) $\lg y = 0,135 - 0,06x + 0,005x^2$
- 5) $\lg y = -0,135 + 0,06x - 0,005x^2$
- 6) $\lg y = -0,54 + 0,24x - 0,02x^2$

г) $y = 1/(a + bx + cx^2)$

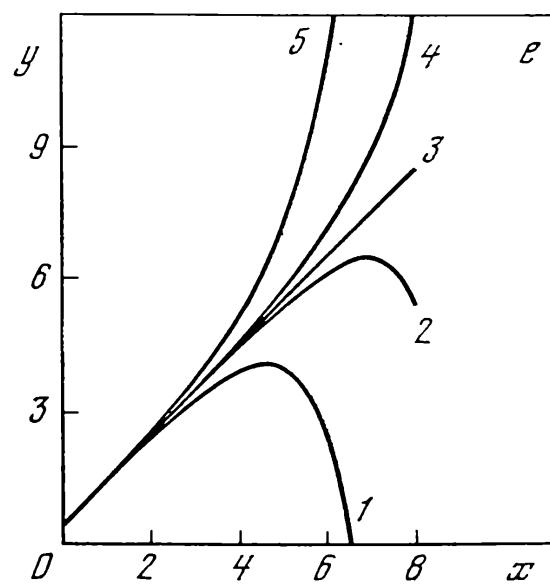
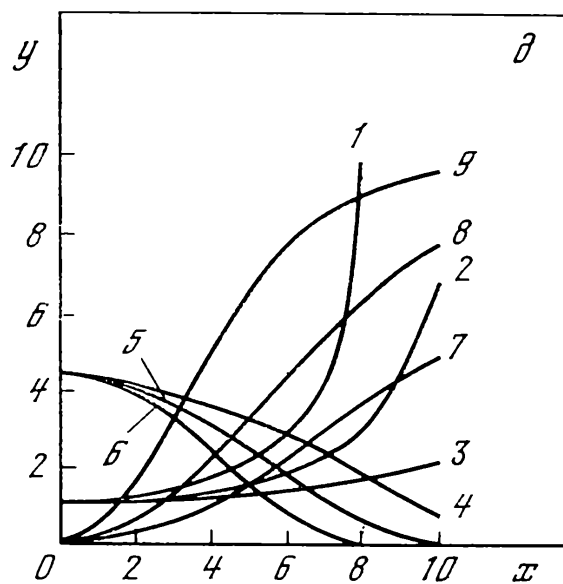
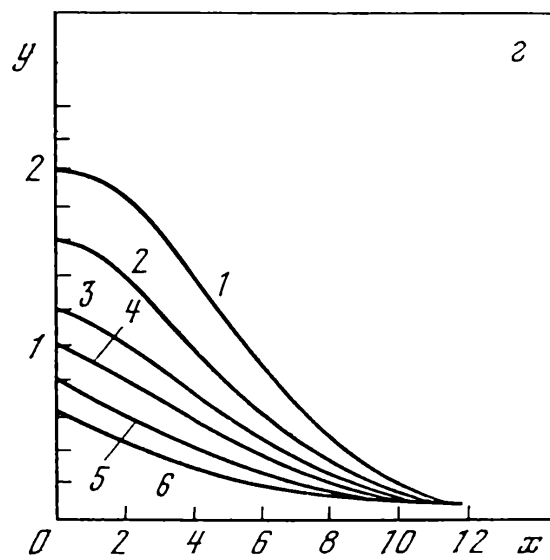
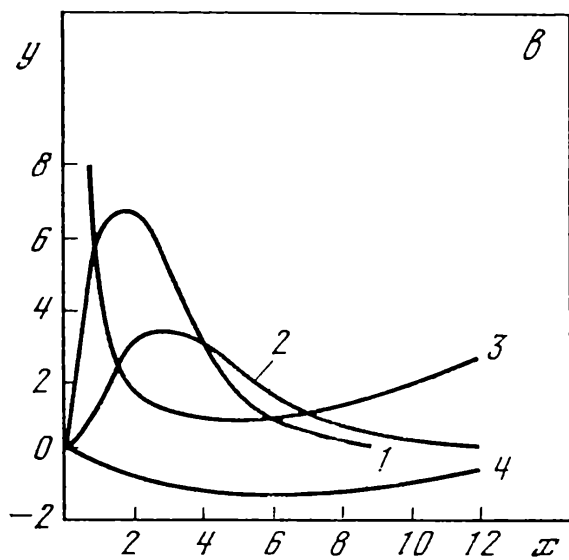
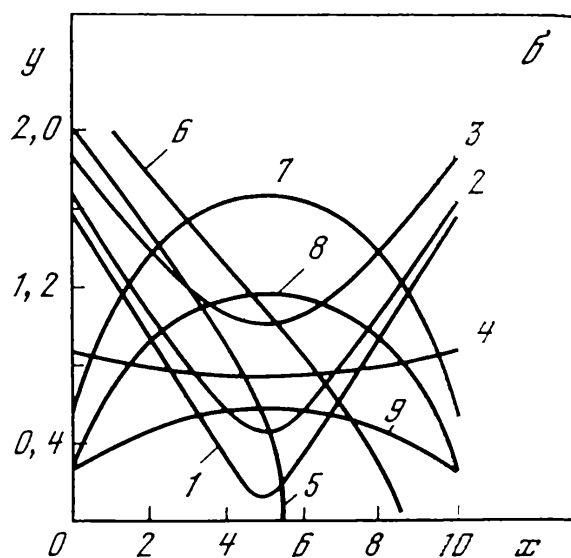
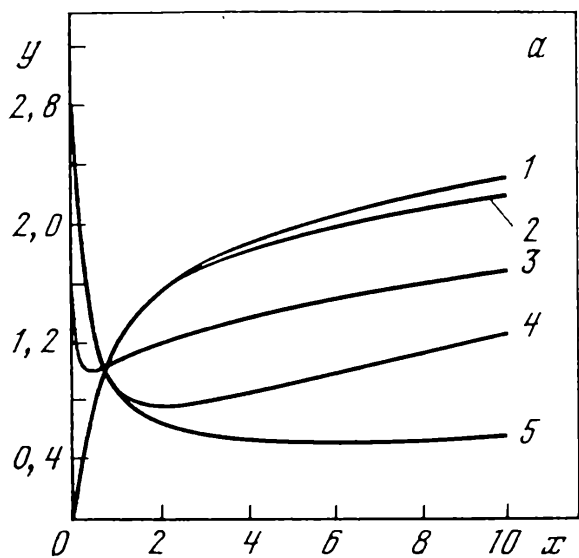
- 1) $y = 1/(0,1x^2 - x + 2,5)$
- 2) $y = 1/(0,1x^2 - x + 2,62)$
- 3) $y = 1/(0,1x^2 - x + 2,7)$
- 4) $y = 1/(0,1x^2 - x + 3)$
- 5) $y = 1/(-0,4x^2 + 4x + 0,1)$
- 6) $y = 1/(-0,1x^2 + x + 0,1)$
- 7) $y = 1/(-0,04x^2 + 0,4x + 0,1)$
- 8) $y = 1/(-0,01x^2 + 0,1x + 0,1)$

д) $y = x/(a + bx + cx^2)$
(см. § 3.05)

- 1) $y = x/(0,2 - 0,1x + 0,05x^2)$
- 2) $y = x/(0,2 - 0,1x + 0,07x^2)$
- 3) $y = x/(0,2 - 0,1x + 0,1x^2)$
- 4) $y = x/(0,2 - 0,1x + 0,2x^2)$
- 5) $y = x/(0,2 - 0,1x + 0,4x^2)$

е) $y = a + \frac{b}{x} + \frac{c}{x^2}$

- 1) $y = 6 - \frac{10}{x} + \frac{5}{x^2}$
- 2) $y = 6 - \frac{10}{x} + \frac{10}{x^2}$
- 3) $y = 6 - \frac{10}{x} + \frac{15}{x^2}$
- 4) $y = 6 - \frac{10}{x} + \frac{20}{x^2}$



теоретические значения дат созревания семян (y') приведены в столбце 5 табл. 3.24. Величины x в этом уравнении соответствуют числам из столбца 3.

Ошибка данного уравнения равна (3.04):

$$m_{y \cdot x} = \sqrt{\frac{\sum (y - y')^2}{N - 3}} = \sqrt{\frac{0,091429}{9 - 6}} = 0,123.$$

На рис. 26 можно видеть, что теоретическая линия регрессии (кривая 2) удовлетворительно совпадает с эмпирическими (кривая 1) данными. Очевидно, что корреляционная связь между сроками зацветания и созревания семян у видов травянистых многолетников уменьшается в конце сезона.

§ 3.14. Метод избранных точек

Существует несколько методов аппроксимации графо-аналитических зависимостей: 1) метод наименьших квадратов; 2) метод моментов; 3) метод средних и 4) метод избранных точек. Два первых метода обеспе-

Рис. 29. Формы кривых третьей группы сложности

а) $y = a + b \lg x + c \lg^2 x$	в) $y = ax^b e^{cx}$ (см. § 3.08)	д) $y = ae^{be^{cx}}$ (см. § 3.10)
1) $y = 1 + \lg x + 0,1 \lg^2 x$ ($y_{\min} = -1,5$ при $\lg x = -5$)	1) $y = 1,5x^{1,5}e^{-0,92x}$	1) $y = e^{0,1e^{0,2x}}$
2) $y = 1 + \lg x + 0,01 \lg^2 x$ ($y_{\min} = -24$ при $\lg x = -50$)	2) $y = 3x^2e^{-0,69x}$	2) $y = e^{0,1e^{0,3x}}$
3) $y = 1 + 0,2 \lg x + 0,3 \lg^2 x$	3) $y = 3x^{-2}e^{0,41x}$	3) $y = e^{0,1e^{0,4x}}$
4) $y = 1 - \lg x + \lg^2 x$	4) $y = -0,5x^{1,5}e^{0,29x}$	4) $y = 5e^{-0,1e^{0,3x}}$
5) $y = 1 - \lg x + 0,5 \lg^2 x$	г) $y = ae^{ax^c}$	5) $y = 5e^{-0,1e^{0,4x}}$
6) $y^2 = a + bx + cx^2$	1) $y = 2,0e^{-0,023x^2}$	6) $y = 5e^{-0,1e^{0,5x}}$
1) $y^2 = 0,1x^2 - x + 2,51$	2) $y = 1,6e^{-0,046x^{1,7}}$	7) $y = 10e^{-5e^{-0,2x}}$
2) $y^2 = 0,1x^2 - x + 2,7$	3) $y = 1,2e^{-0,069x^{1,5}}$	8) $y = 10e^{-5e^{-0,3x}}$
3) $y^2 = 0,1x^2 - x + 3,5$	4) $y = 1,0e^{-0,092x^{1,36}}$	9) $y = 10e^{-5e^{-0,5x}}$
4) $y^2 = 0,01x^2 - 0,1x + 0,3$	5) $y = 0,8e^{-0,115x^{1,24}}$	е) $y = a + bx + ce^{lx}$
5) $y^2 = 4 - x + 0,05x^2$	6) $y = 0,6e^{-0,138x^{1,12}}$	1) $y = 0,5 + x - 0,01e^x$
6) $y^2 = 4,9 - x + 0,05x^2$		2) $y = 0,5 + x - 0,001e^x$
7) $y^2 = 0,25 + x - 0,1x^2$		3) $y = 0,5 + x$
8) $y^2 = 0,05 + 0,5x - 0,05x^2$		4) $y = 0,5 + x + 0,001e^x$
9) $y^2 = 0,05 + 0,1x - 0,01x^2$		5) $y = 0,5 + x + 0,01e^x$

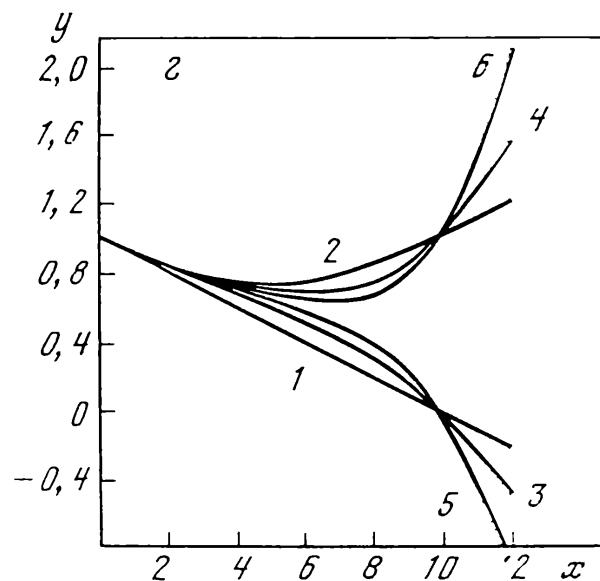
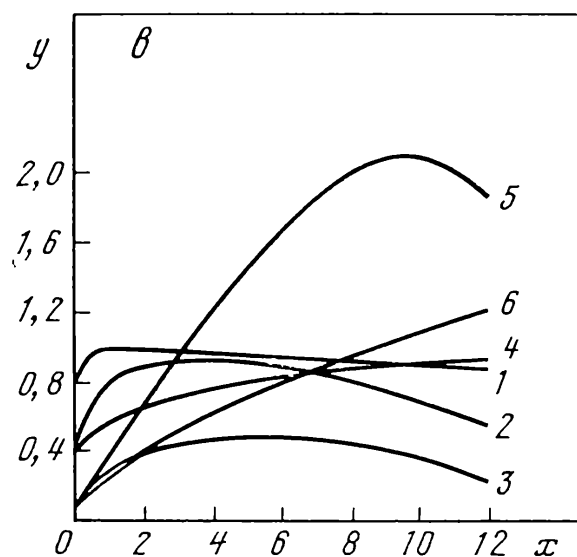
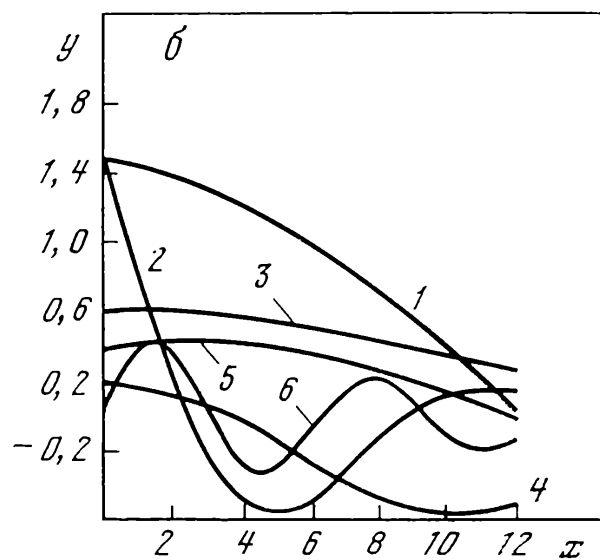
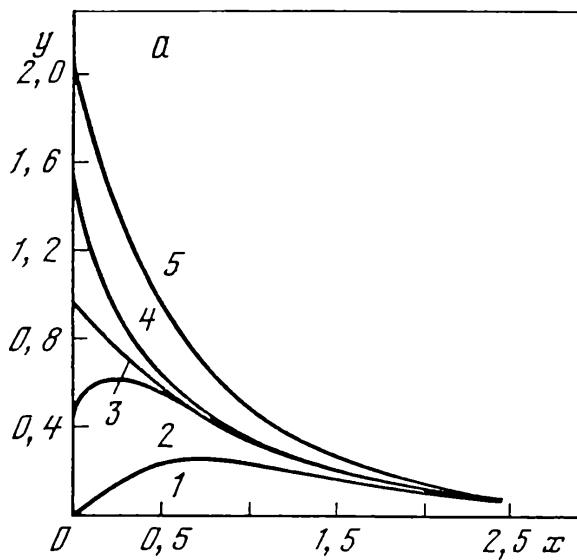


Рис. 30. Формы кривых четвертой группы сложности

а) $y = ae^{bx} + ce^{dx}$

- 1) $y = e^{-x} - e^{-2x}$
- 2) $y = e^{-x} - 0,5e^{-5x}$
- 3) $y = e^{-x}$
- 4) $y = e^{-x} + 0,5e^{-5x}$
- 5) $y = e^{-x} + e^{-2x}$

в) $y = ax^b + cx^d$

- 1) $y = 2x^{0,1} - x^{0,2}$
- 2) $y = 3x^{0,5} - 2,2x^{0,6}$
- 3) $y = 2,3x^{0,8} - 2x^{0,85}$
- 4) $y = 0,1x^{0,1} + 0,5x^{0,2}$
- 5) $y = 0,33x - 0,0012x^3$
- 6) $y = 0,25x^{0,5} + 0,05x^{0,8}$

б) $y = e^{ax} (c \cos bx + d \sin bx)$

- 1) $y = e^{0,01x} (1,5 \cos 0,1x - 0,5 \sin 0,1x)$
- 2) $y = e^{-0,2x} (1,5 \cos 0,5x - 0,5 \sin 0,5x)$
- 3) $y = e^{-0,1x} (0,6 \cos 0,1x + 0,8 \sin 0,1x)$
- 4) $y = e^{0,1x} (0,2 \cos 0,3x - 0,1 \sin 0,3x)$
- 5) $y = e^{0,02x} (0,4 \cos 0,16x + 0,17 \sin 0,16x)$
- 6) $y = 0,5e^{-0,1x} \sin x$

г) $y = a + bx + cx^2 + dx^3 \dots$ (см. § 3.04, 3.06)

- 1) $y = 1 - 0,1x$
- 2) $y = 1 - 0,1x + 0,01x^2$
- 3) $y = 1 - 0,1x + 0,01x^2 - 0,001x^3$
- 4) $y = 1 - 0,1x + 0,01x^2 - 0,001x^3 + 0,0001x^4$
- 5) $y = 1 - 0,1x + 0,01x^2 - 0,001x^3 + 0,0001x^4 - 0,00001x^5$
- 6) $y = 1 - 0,1x + 0,01x^2 - 0,001x^3 + 0,0001x^4 - 0,00001x^5 + 0,000001x^6$

печивают практически любую требуемую точность аппроксимации, однако применимы далеко не ко всем типам зависимостей из-за отсутствия способов аналитического решения некоторых видов уравнений и их систем. Оба первых метода также довольно трудоемки в отношении вычислительной работы. Третий метод занимает промежуточное положение как по точности аппроксимации, так и по объему вычислений между первыми двумя и четвертым методом. Заметим, однако, что благодаря подробно разработанной методике (Семендяев, 1933) метод средних более универсален по сфере приложения, чем методы моментов и наименьших квадратов.

Метод избранных точек наиболее прост по технике вычислительной работы, но в общем случае не гарантирует особой точности. Между тем с появлением и увеличивающимся распространением ЭВМ с программируемым процессом вычислений появились новые возможности использования уже известных вычислительных методов.

Дело в том, что эффективность, или рентабельность, применения ЭВМ складывается в том числе из соотношения между работой по составлению алгоритма и программированию, с одной стороны, и работой по непосредственному машинному счету, с другой стороны. Чем меньше это отношение, тем эффективнее применяется ЭВМ при прочих равных условиях.

Иначе говоря, выгоднее меньше тратить времени и сил на составление алгоритма, если есть возможность адекватно переложить значительную долю работы на машину, конечно при равной полноте и точности результатов в обоих случаях. Обычно достичь указанной ситуации удастся, упрощая алгоритм за счет увеличения числа вычислительных операций, производимых ЭВМ. Таковы, например, некоторые вычислительные процедуры, связанные с применением итераций.

В отношении к аппроксимации зависимостей итерации часто применяются в таких процедурах, где коэффициенты уравнений функций находятся способом последовательных улучшений, причем критерием оценки качества приближения величин коэффициентов служит обычно дисперсия значений функции или величина ошибки последней. Как только при последовательном уменьшении или увеличении значения первого коэффициента дисперсия функции перестает уменьшаться, переходят поочередно таким же образом к следующим коэффициентам, а затем процедура может повториться нужное число раз до тех пор, когда при начале увеличения или уменьшения каждого коэффициента дисперсия функции будет только увеличиваться. На этом вычисления коэффициентов заканчиваются. Недостаток описанного способа итераций заключается в том, что исходные значения коэффициентов могут быть заданы с большими отклонениями и потребуется слишком много машинного времени для их уточнения.

Сократить время расчетов при этом можно двумя способами: оперативно изменяя шаг итераций или задавая в качестве исходных величин приближенно найденные, например по способу избранных точек значения коэффициентов уравнения. В последнем случае весь-

ма полезны модельные графики кривых, построенные по уравнениям с различными величинами коэффициентов и позволяющие понять закон изменения формы кривой в зависимости от изменения величины коэффициентов ее уравнения.

Наиболее полный сборник таких графиков различных кривых, составленный К. А. Семендяевым (1933), в настоящее время стал библиографической редкостью, поэтому в данном параграфе приводятся модельные графики (рис. 27—30) из этого издания. Сравнивая эмпирическую кривую исследуемого явления с этими графиками, можно подобрать тип кривой и подходящее уравнение для ее аппроксимации методами избранных точек или итераций. Определение коэффициентов уравнений некоторых из функций, показанных на рис. 27—30, можно сделать также по способам, приведенным в параграфах, номера которых указаны в подписях к этим рисункам. Рис. 27—30 и формулы соответствующих им кривых для удобства подбора подходящего уравнения расположены по группам сложности расчетов: от менее к более сложным.

СРАВНЕНИЕ ВЫБОРОК И СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ

§ 4.01. Сравнение средних арифметических и дисперсий

Установление статистической достоверности различий между результатами опытов является наиболее распространенной, типичной задачей в экспериментальной ботанике и агрономии. Если требуется количественно сравнить два взвешенных или невзвешенных ряда дат по нижеописываемому критерию Стьюдента, то необходимо знать: 1) их средние арифметические (см. § 1.03—1.04); 2) дисперсии или средние квадратические отклонения (см. §§ 1.09, 1.11), или ошибки средних арифметических (см. § 1.03); 3) объемы сравниваемых выборок; 4) тип распределения вариантов, который должен быть хотя бы приблизительно нормальным.

Сравнивать совокупности можно по их средним арифметическим и по их вариабельности. Рекомендуются это делать одновременно. Методы сравнения больших и малых выборок различаются между собой.

I. Малые выборки ($N \leq 20$).

Когда хотя бы одна из сравниваемых выборок мала, то поступаем следующим образом. Сначала проверяем гипотезу о том, что дисперсии обоих рядов равны. В зависимости от того, достоверным или недостоверным является различие дисперсий, последующее сравнение средних арифметических малых выборок производят двумя способами. Рассмотрим два примера сравнения средних арифметических малых выборок с равными и неравными дисперсиями.

Пример 1. Сравнивались средние арифметические диаметров соцветий у двух сортов нивяника. У одного сорта средний диаметр соцветия был: $M_1 = 5$ см, дисперсия $\sigma_1^2 = 2,5$, объем выборки $N_1 = 10$; у другого сорта: $M_2 = 7$ см, $\sigma_2^2 = 2,4$, $N_2 = 14$. Сравнение дисперсий производится следующим образом:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{2,5}{2,4} = 1,04. \quad (4.01)$$

В качестве числителя в этом отношении всегда берется бо́льшая дисперсия. Находим для каждой дисперсии число степеней свободы, которое равно объему выборки без единицы, т. е.:

$$\nu(1) = N_1 - 1 = 10 - 1 = 9; \quad \nu(2) = N_2 - 1 = 14 - 1 = 13.$$

По табл. 9П при $P_1 = 95\%$ на пересечении столбца 8 (вместо 9, которого нет в таблице) и строки 13: $F'(95\%) = 2,77$, что больше $F = 1,04$; следовательно, по колеблемости диаметра соцветий два

данных сорта достоверно не различаются. В случае несущественно различающихся или равных дисперсий средние арифметические сравниваются по формуле

$$|t| = (M_1 - M_2) \sqrt{\frac{1 - 2(N_1 + N_2)^{-1}}{\frac{\sigma_1^2}{N_2} + \frac{\sigma_2^2}{N_1}}}, \quad (4.02)$$

где t — критерий Стьюдента; M_1, M_2 — сравниваемые средние арифметические; σ_1^2, σ_2^2 — дисперсии сравниваемых выборок; N_1, N_2 — их объемы. Подставляя в формулу (4.02) числовые значения, получим:

$$|t| = (5 - 7) \sqrt{\frac{1 - 2(10 + 14)^{-1}}{\frac{2,5}{14} + \frac{2,4}{10}}} = 2,96.$$

Знак критерия не принимается во внимание.

По табл. 3П при числе степеней свободы $\nu = N_1 + N_2 - 2 = 22$ и $P'_1 = 0,95$ $t = 2,074$, что меньше вычисленного значения t ; следовательно, данные два сорта нивяника достоверно различаются между собой по диаметру соцветия.

Пример 2. В § 4.07 приведены данные по длине колоса у двух сортов пшеницы. Сравним их при помощи критерия Стьюдента. У сорта А: $M_1 = 9,2$; $\sigma_1^2 = 0,7$; $\sigma_1 = 0,8$; $m_1 = 0,3$; $N_1 = 10$; у сорта Б: $M_2 = 7,9$; $\sigma_1^2 = 0,1$; $\sigma_2 = 0,4$; $m_2 = 0,1$; $N_2 = 9$.

Сравним дисперсии: $F = \sigma_1^2 / \sigma_2^2 = 0,7 / 0,1 = 7$. По табл. 9П при доверительном уровне $P_1 = 95\%$ и числах степеней свободы $\nu(1) = N_1 - 1 = 10 - 1 = 9$; $\nu(2) = N_2 - 1 = 9 - 1 = 8$ находим: $F'(95) = 3,44$. Следовательно, имеется существенное различие по колеблемости длины колоса у данных сортов. При различных по величине дисперсиях разница средних арифметических оценивается по формуле

$$|t| = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}, \text{ или } |t| = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \quad (4.03)$$

с числом степеней свободы:

$$\nu = (N_1 + N_2 - 2) \left[\frac{1}{2} + \left(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} \right)^{-1} \right]. \quad (4.04)$$

Подставляя имеющиеся данные в формулы (4.03) и (4.04), где m_1 и m_2 — ошибки средних, а остальные обозначения те же, что и в (4.02), получим:

$$t = \frac{9,2 - 7,9}{\sqrt{0,3^2 + 0,1^2}} = 4,11,$$

$$\nu = (10 + 9 - 2) \left(\frac{1}{2} + \frac{1}{\frac{0,7}{0,1} + \frac{0,1}{0,7}} \right) = 10,9.$$

Число степеней свободы округляем до целого числа $\nu = 11$. По табл. 3П при доверительном уровне $P_1 = 95\%$ $t = 2,201$, что меньше вычисленного критерия: $2,201 < 4,11$. Следовательно, между данными сортами пшеницы имеется достоверная разница в средней длине их колосьев.

II. Большие выборки ($N_1 > 20$; $N_2 > 20$).

Когда обе сравниваемые выборки существенно велики, а их объемы незначительно различаются или равны, сравнение средних производится по той же формуле (4.03):

$$|t| = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}},$$

с числом степеней свободы: $\nu = N_1 + N_2 - 2$; где m_1, m_2 — ошибки сравниваемых средних арифметических M_1 и M_2 с объемами выборок N_1 и N_2 .

Рассмотрим пример вычисления критерия Стьюдента в этом случае. Сравнивались фенодаты распускания почек у группы одних и тех же 57 видов деревьев и кустарников, за одни и те же годы в Москве и Ленинграде. Статистики обоих массивов наблюдений оказались следующими соответственно: $M_1 = 57,9$; $\sigma_1 = 19,1$; $N_1 = 57$; $M_2 = 65,8$; $\sigma_2 = 22,0$; $N_2 = 57$, где M_1 и M_2 — даты распускания почек, в днях от 1 марта (см. табл. 24II).

Подставляя эти числа в формулу (4.03), получим:

$$|t| = \frac{57,9 - 65,8}{\sqrt{\frac{19,1^2}{57} + \frac{22,0^2}{57}}} = 2,048, \quad \nu = 57 + 57 - 2 = 112.$$

По табл. 3П при ближайшем меньшем числе степеней свободы 100, на 95 % доверительном уровне $t = 1,982$, что меньше вычисленной величины. Следовательно, древесные растения в Москве распускают почки примерно на неделю раньше, чем в Ленинграде.

Если сравниваемые совокупности достаточно велики по объему и распределение их величин подчиняется нормальному закону, то сравнить средние можно без помощи таблиц и не прибегая к извлечению корней по формуле:

$$\frac{(M_1 - M_2)^2}{m_1^2 + m_2^2} \geq 9, \quad (4.05)$$

где M_1, M_2 — сравниваемые средние, а m_1, m_2 — их ошибки. Если левая часть неравенства будет больше 9, то различие средних достоверно значимо, а если она будет меньше 9, то различие средних нельзя считать достоверным.

§ 4.02. Парный критерий Стьюдента

Выборки с попарно связанными вариантами можно сравнивать также и по методам из предыдущего параграфа. Однако ошибка разности средних в этом случае будет больше из-за того, что не учитывается попарная связь вариантов и, следовательно, вычисленная величина

критерия Стьюдента будет занижена. Но, с другой стороны, по приводимому ниже методу число степеней свободы берется в 2 раза меньше, чем по методам из § 4.01, и реально существующее различие может быть не обнаружено из-за большой табличной величины критерия. Поэтому в сомнительных случаях следует вычислить коэффициент корреляции или иной показатель связи между сравниваемыми попарными рядами. Если корреляция между ними окажется существенной, то лучше применить метод из § 4.02, а если связи не окажется, то методы — из § 4.01.

В табл. 4.01 приведены фенодаты начала цветения одной и той же группы из 12 видов деревьев и кустарников (в днях) от 1 марта (см. табл. 24П) за два года — 1949 и 1958. Заметно, что в 1958 г. зацветание у большинства видов наступило позднее, чем в 1949 г. Требуется оценить достоверность такого предположения. Рекомендуется следующая последовательность вычислений.

1. Показатель корреляции рангов Спирмена оказался достоверным и равным $\rho = 0,745$. Следовательно, целесообразно применить парный критерий Стьюдента, а не методы из § 4.01.

2. Находим средние арифметические столбцов 1 и 2, они соответственно равны: $M_1 = 114,58$; $M_2 = 127,42$.

3. В столбце 3 записаны попарные разности чисел из столбцов 1 и 2.

4. В столбце 4 вычислены квадраты чисел из столбца 3.

5. Суммируем числа столбцов 3 и 4:

$$\Sigma \delta = 154; \quad \Sigma \delta^2 = 3330.$$

6. Делаем проверку вычислений по равенству

$$|M_1 - M_2| = \frac{\Sigma \delta}{N},$$

$$114,58 - 127,42 = 154/12; \quad 12,8 = 12,8.$$

Вычисления сделаны верно.

7. Находим величину критерия по формуле

$$t = \frac{\Sigma \delta}{\sqrt{\frac{N \Sigma \delta^2 - (\Sigma \delta)^2}{N - 1}}}, \quad (4.06)$$

$$t = \frac{154}{\sqrt{\frac{12 \cdot 3330 - (154)^2}{11}}} = 4,01.$$

При числе степеней свободы $\nu = N - 1 = 12 - 1 = 11$, на 95 % уровне достоверности по табл. 3П, $t = 2,201$. Вычисленное его значение больше табличного, поэтому делаем вывод о достоверности различий в сроках начала цветения данной группы видов в 1949 и 1958 гг.

Если бы для проверки высказанного предположения мы воспользовались формулой (4.02), которую здесь следовало бы применить

Таблица 4.01. Сравнение дат зацветания 12 видов за два года

Даты зацветания, в днях от 1 марта		(2-1)= δ	δ^2
1949 г.	1958 г.		
112	136	24	576
120	149	29	841
127	136	9	81
102	117	15	225
149	158	9	81
133	117	-16	256
128	142	14	196
127	137	10	100
133	142	9	81
61	81	20	400
99	117	18	324
84	97	13	169
1375	1529	154	3330

Таблица 4.02. Урожай двух сортов нута (в ц/га)

Год	Урожай сорта 1	Урожай сорта 2
1956	28,0	25,7
1957	21,3	12,2
1958	30,0	15,0
1959	15,0	18,0
1960	22,0	16,7
1961	18,1	17,1
1962	16,6	17,0
1963	18,0	16,0
1964	22,1	19,2
1965	22,0	20,0
N=10	213,1	176,9

ввиду несущественного различия дисперсий сравниваемых рядов ($F = 1,21$), то оказалось бы, что различия между годами в фенодатах нет: $t = 1,28$ против табличного $t(0,95) = 2,07$. Отсюда видно, что неправильно выбранный метод сравнения средних может привести к неточности и даже к ошибочному противоположному заключению при оценке гипотезы.

§ 4.03. Сравнение коэффициентов вариации и показателей точности опыта

В сельском хозяйстве важным показателем сорта является стабильность его урожаев. Так, более целесообразно культивировать сорт со средней, но постоянной урожайностью по сравнению с высокопродуктивным сортом, урожай которого, однако, сильно колеблется по годам. Выводенность сорта по тем или иным декоративным признакам имеет большое значение в цветоводстве, а также для механизации работ по возделыванию различных полезных растений и уборке урожая, а также для промышленной обработки сырья из них. Во всех подобных случаях возникает необходимость сравнить две выборки по степени их варьирования в относительных числах. Это можно сделать по формуле:

$$\frac{|v_1 - v_2|}{\sqrt{m_{v1}^2 + m_{v2}^2}} \leq 3 + \frac{6}{N-4}, \quad (4.07)$$

где v_1, v_2 — коэффициенты вариации сравниваемых рядов; m_{v1}, m_{v2} — ошибки сравниваемых коэффициентов вариации; N — объем меньшей из сравниваемых выборок.

Если левая часть неравенства (4.07) будет больше правой части, то различие коэффициентов вариации считается достоверным. При пользовании формулой (4.07) объем меньшей из выборок должен быть не меньше 5.

Сравним два сорта нута (табл. 4.02) с целью установить, имеет ли сорт 2 преимущество перед сортом 1 в отношении стабильности урожая.

Сигмы и средние арифметические для этих двух сортов соответственно имеют значения:

$$\sigma_1 = 4,77, \quad M_1 = 21,31 \pm 1,51;$$

$$\sigma_2 = 3,56 \quad M_2 = 17,69 \pm 1,13,$$

откуда коэффициенты вариации и их ошибки по формулам (1.43) и (1.47) равны:

$$v_1 = 22,40\%; \quad v_2 = 20,11\%; \quad m_{v1} = 5,25\%; \quad m_{v2} = 4,67\%:$$

Подставляя найденные величины в формулу (4.07):

$$\frac{22,40 - 20,11}{\sqrt{5,25^2 - 4,67^2}} \geq 3 + \frac{6}{10 - 4},$$

получим: $0,3 < 4$. Поскольку левая часть неравенства меньше правой, различие двух сравниваемых сортов нута по варьированию их урожаев считаем несущественным.

Сравнение показателей точности опыта производится примерно в тех же случаях, что и сравнение коэффициентов вариации, поэтому может заменить последнее или служить контролем к нему. В дисперсионном анализе при обработке данных полевых опытов, где принято использовать показатель точности опыта для оценки качества проведения экспериментальных работ, сравнение показателей точности опыта с этой целью может иметь свое самостоятельное применение.

Два сорта нута (табл. 4.02) по формулам (1.50), (1.52) имеют следующие показатели точности опыта и их ошибки:

$$P_1 = 7,08; \quad P_2 = 6,36; \quad m_{P1} = 1,66; \quad m_{P2} = 1,48.$$

По аналогии с неравенством (4.07) показатели точности опыта сравниваются по формуле:

$$\frac{|P_1 - P_2|}{\sqrt{m_{P1}^2 + m_{P2}^2}} \leq 3 + \frac{6}{N - 4}, \quad (4.08)$$

где P_1, P_2 — показатели точности опыта сравниваемых выборок, m_{P1}, m_{P2} — ошибки этих показателей; N — объем меньшей из сравниваемых выборок.

По формуле (4.08):

$$\frac{|7,08 - 6,36|}{\sqrt{1,66^2 + 1,48^2}} \leq 3 + \frac{6}{N - 4}, \quad \text{откуда } 0,3 < 4.$$

Следовательно, два сравниваемых сорта нута достоверно не различаются по величине показателей точности опыта и, таким образом, подтверждается тот же вывод, сделанный при сравнении коэффициентов вариации.

§ 4.04. Сравнение долей из больших выборок

Метод применяется в том случае, когда выборки, проценты или доли из которых сравниваются, достаточно велики по объему и принадлежат к различным совокупностям. Приближенным критерием достаточной величины выборки в данном случае может служить неравенство: $pN \geq 500$, т. е. произведение процента (p) особей, обладающих изучаемым признаком в выборке, на объем этой выборки должно быть 500 или больше.

Сравнение долей производится по формуле:

$$t = \frac{p_1 - p_2}{\sqrt{p'q' \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}, \quad (4.09)$$

где средняя долей:

$$p' = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}, \quad q' = 100 - p'.$$

Имея в виду, что

$$p_1 = \frac{n_1 \cdot 100}{N_1}, \quad p_2 = \frac{n_2 \cdot 100}{N_2},$$

усреднить доли можно также по более простой формуле:

$$p' = \frac{n_1 + n_2}{N_1 + N_2} \cdot 100, \quad (4.10)$$

где n_1, n_2 — численности особей с наблюдаемым признаком в обеих выборках; p_1, p_2 — сравниваемые доли особей с наблюдаемым признаком, в процентах; N_1, N_2 — объемы сравниваемых выборок; p' — среднее значение долей, в процентах; q' — дополнение к p' , выражающее среднюю долю особей, не обладающих изучаемым признаком, в процентах; t — критерий Стьюдента, для трех стандартных уровней достоверности он равен 1,96; 2,58; 3,29.

Следовательно, при сравнении долей из больших выборок прибегать к табл. 3П нет необходимости.

Рассмотрим пример. Изучались особенности заражения фомозом корнеплодов моркови. При естественном заражении фомозом из $N_1 = 1098$ корнеплодов заболело 82, или $p_1 = 7\%$, при искусственном заражении заболело 639 корнеплодов из $N_2 = 1008$, т. е. $p_2 = 63\%$. Подставляя эти данные в формулу (4.09), получим:

$$p' = \frac{1098 \cdot 7 + 1008 \cdot 63}{1098 + 1008} = 34\%,$$

$$q' = 100 - 34 = 66\%;$$

$$|t| = \frac{7 - 63}{\sqrt{34 \cdot 66 (1/1098 + 1/1008)}} = 27,2,$$

что больше любого из трех стандартных уровней критерия, приведенных выше. Следовательно, доля больных корнеплодов в том варианте опыта, где они заражались искусственно, существенно и достоверно отличается от доли больных корнеплодов в варианте с естественным их заражением.

Сравнение долей из двух различных совокупностей можно производить также по критерию хи-квадрат из § 4.14, расположив данные в виде четырехпольной таблицы, причем доли не требуется переводить в проценты, взяв вместо них непосредственно первичные данные.

§ 4.05. Сравнение долей по методу Фишера

При помощи этого метода можно сравнивать доли как из больших, так и малых выборок, причем с одной и той же точностью для тех и других, что обеспечивается предварительным преобразованием долей в радианы (угловая мера) по формуле $\varphi = 2 \arcsin \sqrt{p}$ (табл. 15П). Сравнение проводится затем посредством критерия Фишера по формуле:

$$F(\varphi_1 - \varphi_2)^2 \frac{N_1 N_2}{N_1 + N_2}, \quad (4.11)$$

где F — критерий Фишера; φ_1, φ_2 — числа, полученные в результате преобразования долей по табл. 15П; N_1, N_2 — объемы сравниваемых выборок.

Рассмотрим применение метода Фишера на примере из предыдущего параграфа. Сравниваемые доли равны: $p_1 = 7\% = 0,07$, $p_2 = 63\% = 0,63$; объемы выборок: $N_1 = 1098$, $N_2 = 1008$. По табл. 15П находим, что p_1 и p_2 соответствуют: $\varphi_1 = 0,536$, $\varphi_2 = 1,834$. Подставляя эти числа в формулу (4.11), получим:

$$F = (0,536 - 1,834)^2 \frac{1098 \cdot 1008}{1098 + 1008} = 885.$$

При пользовании данным методом первое число степеней свободы всегда равно единице: $\nu(1) = 1$, а $\nu(2) = N_1 + N_2 - 2 = 1098 + 1008 - 2 = 2104$.

По табл. 9П на уровне 95 % находим $F' = 3,84$. Вычисленное число $F = 885$ значительно больше табличного, поэтому заключаем, что сравниваемые совокупности вполне достоверно различаются по доле больных растений.

Оценку достоверности различия долей можно сделать также и посредством критерия Стьюдента по формуле

$$t = (\varphi_1 - \varphi_2) \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \geq 1,96 (2,58), \quad (4.12)$$

где обозначения те же, что и к формуле (4.11).

В случаях, требующих особой точности, желательно ввести поправку на непрерывность к сравниваемым долям.

Рассмотрим на том же примере применение критерия Стьюдента для оценки достоверности различия долей с введением поправки на непрерывность к последним. По данным § 4.04 объемы сравниваемых выборок: $N_1 = 1008$ и $N_2 = 1098$ корнеплодов, число больных корнеплодов в них: $n_1 = 639$, $n_2 = 82$, откуда доли равны:

$$p_1 = 639/1008 = 0,63393, \quad p_2 = 82/1098 = 0,074681.$$

При введении поправки на непрерывность по следующим формулам p_1 должно быть в них больше p_2 :

$$\begin{aligned} p'_1 &= p_1 - \frac{0,5}{N_1} = 0,63393 - \frac{0,5}{1008} = 0,63343; \\ p'_2 &= p_2 + \frac{0,5}{N_2} = 0,074681 + \frac{0,5}{1098} = 0,075136. \end{aligned} \quad (4.13)$$

После преобразования исправленных величин долей в радианы (табл. 15П) оцениваем значимость их разницы по критерию Стьюдента:

$$t = (1,841 - 0,555) \sqrt{\frac{1008 \cdot 1098}{1008 + 1098}} = 29,5.$$

Полученная величина критерия больше обоих его критических значений: 1,96 на 95 % доверительном уровне и 2,58 на 99 % доверительном уровне. Поэтому считаем, что сравниваемые совокупности существенно различаются по доле больных растений в них.

§ 4.06. Сравнение коэффициентов корреляции

Когда вычислены коэффициенты корреляции по некоторым близким между собой данным, может возникнуть вопрос о достоверности их различия. Если между коэффициентами корреляции нет достоверной разницы, то при качественно одинаковых объектах сравнения можно заключить, что обе выборки по данному признаку относятся к одной генеральной совокупности.

При изучении двух популяций гелениума осеннего в одной из них коэффициент корреляции между числом стеблей и числом соцветий на особи был равен $r_1 = 0,67$; в другой та же связь выражалась через $r_2 = 0,61$. Численности выборок были: $N_1 = 69$ и $N_2 = 45$ растений. Относятся ли обе популяции по этому признаку к одной совокупности, или иначе говоря, одинаковы ли коэффициенты корреляции между числом стеблей и числом соцветий у обеих популяций? Для ответа на этот вопрос требуется выполнить следующие действия.

1. По табл. 26П переводим r_1 и r_2 в величины $z_1 = 0,8107$ и $z_2 = 0,7089$.

2. Оцениваем разницу между z_1 и z_2 критерием Стьюдента:

$$t = \frac{|z_2 - z_1|}{\sqrt{\frac{N_1 + N_2 - 6}{(N_1 - 3)(N_2 - 3)}}} = \frac{|0,7089 - 0,8107|}{\sqrt{\frac{69 + 45 - 6}{(69 - 3)(45 - 3)}}} = 0,516. \quad (4.14)$$

3. Число степеней свободы при оценке разницы коэффициентов корреляции принимается равным:

$$\nu = N_1 + N_2 - 4 = 69 + 45 - 4 = 110.$$

По табл. 3П на уровне значимости $0,05 \ t' = 1,98$, что больше вычисленного значения, равного 0,516. Следовательно, между двумя изучаемыми популяциями гелениума осеннего нет достоверного различия, и их можно относить к одной совокупности при изучении указанных взаимосвязей.

§ 4.07. Критерий лямбда

Критерии различия, при помощи которых могут быть сравнены статистические совокупности, разделяются на две группы: параметрические (§ 4.01—4.06) и непараметрические (§ 4.07—4.15). К первой группе относятся критерии, для применения которых необходимо вычислить среднюю арифметическую, сигму или ошибки параметров. Непараметрические критерии не требуют для своего применения вычисления названных показателей, что упрощает процесс сравнения совокупностей. Критерий лямбда относится к непараметрическим критериям сравнения и, в отличие от многих из них, не требует специальных таблиц для оценки достоверности различия. Критерий лямбда вычисляется по формуле

$$\lambda^2 = \delta_{\max}^2 \frac{N_x N_y}{N_x + N_y}, \quad (4.15)$$

где λ — критерий различия лямбда; δ_{\max} — максимальная из попарных разностей чисел, приведенных в столбцах 6 и 7 табл. 4.03; N_x, N_y — объемы сравниваемых рядов.

Если $N_x = N_y$, то λ^2 вычисляется по более простой формуле (4.16). Если вычисленное значение λ^2 превышает число 1,84 (при 5 % уровне значимости) или число 2,65 (при 1 % уровне значимости), то различие между сравниваемыми совокупностями достоверно.

Сравним два сорта пшеницы по длине стержня их колоса. Было измерено $N_x = 10$ колосьев сорта А (в см): 9,0; 9,0; 9,0; 10,5; 9,7; 10,4; 9,3; 7,9; 8,4; 8,5; и $N_y = 9$ колосьев сорта Б: 8,5; 7,5; 8,0; 8,0; 8,2; 7,4; 8,2; 7,9; 7,7. Действия по вычислению критерия различия лямбда производим в следующем порядке по столбцам табл. 4.03.

1. Ранжируем варианты рядов x и y в один ряд, в возрастающем или нисходящем порядке их величин.

2. Отмечаем единицей частоты, относящиеся к ряду x , и нулем — частоты ряда y .

3. Отмечаем единицей частоты, относящиеся к ряду y , и нулем — частоты ряда x . Числа в столбцах 2, 3, попарно совпадающие по величине, исключаем из расчетов.

4. Составляем накопленные частоты ряда x , т. е. складываем частоты нарастающим итогом. Сумма накопленных частот в конце столбца 4 должна равняться объему выборки N_x и итогу столбца 2.

Таблица 4.03. Вычисление максимальной разности и накопленных частот для применения критерия лямбда

x, y	f_x	f_y	Σf_x	Σf_y	$(\Sigma f_x)/N_x$	$(\Sigma f_y)/N_y$	$ \delta $
1	2	3	4	5	6	7	8
7,4	0	1	0	1	0,000	0,111	0,111
7,5	0	1	0	2	0,000	0,222	0,222
7,7	0	1	0	3	0,000	0,333	0,333
7,9*	1*	0	1	3	0,100	0,333	0,233
7,9	0	1	1	4	0,100	0,444	0,344
8,0	0	1	1	5	0,100	0,556	0,456
8,0	0	1	1	6	0,100	0,667	0,567
8,2	0	1	1	7	0,100	0,778	0,678
8,2*	0	1	1	8	0,100	0,889	0,789
8,4	1	0	2	8	0,200	0,889	0,689
8,5	1	0	3	8	0,300	0,889	0,589
8,5	0	1	3	9	0,300	1,000	0,700
9,0	1	0	4	9	0,400	1,000	0,600
9,0	1	0	5	9	0,500	1,000	0,500
9,0	1	0	6	9	0,600	1,000	0,400
9,3	1	0	7	9	0,700	1,000	0,300
9,7	1	0	8	9	0,800	1,000	0,200
10,4	1	0	9	9	0,900	1,000	0,100
10,5	1	0	10	9	1,000	1,000	0,000

$N_x=10$ $N_y=9$

5. Составляем накопленные частоты ряда y . Сумма накопленных частот в конце столбца 5 должна равняться объему выборки N_y и итогу столбца 3.

6. Числа столбца 4 поочередно делим на объем ряда x , т. е. на 10, и записываем в столбце 6.

7. Числа столбца 5 делим на объем ряда y , т. е. на 9, и записываем в столбце 7.

8. Вычисляем без учета знака попарные разности чисел столбцов 6 и 7, выбираем максимальную разность, в данном случае она равна $\delta_{max} = 0,789$.

По формуле (4.15) получаем:

$$\lambda^2 = 0,789^2 \cdot \frac{10 \cdot 9}{10 + 9} = 2,95.$$

Вычисленная величина больше вышеприведенных обоих критических значений:

$$2,95 > 2,65 > 1,84,$$

поэтому делаем вывод о том, что два сравниваемых сорта пшеницы достоверно различаются по длине стержня колоса.

При $N_x = N_y$ расчет критерия лямбда проводится по более простой формуле:

$$\lambda^2 = \frac{\delta_{\max}^2}{2N}, \quad (4.16)$$

где λ^2 — квадрат критерия лямбда; δ_{\max} — максимальная разность накопленных частот из столбцов 4 и 5 табл. 4.03; N — объем одного из сравниваемых рядов. Таким образом, при одинаковом объеме сравниваемых рядов отпадает необходимость в столбцах 6 и 7.

Пример. Если удалить из ряда x число 7,9* (столбец 1) и его частоту 1* (столбец 2), то столбцы 4 и 5 табл. 4.03 будут выглядеть следующим образом (в горизонтальном расположении):

4	$\sum f_x$	0	0	0	0	0	0	0	0*	1	2	2	3	4	5	6	7	8	9
5	$\sum f_y$	1	2	3	4	5	6	7	8*	8	8	9	9	9	9	9	9	9	9

Максимальная разность между этими числами при попарном сравнении остается на том же месте, против значения 8,2* и равна: $8* - 0* = 8$. Отсюда по формуле (4.16): $\lambda^2 = 8^2/(2 \cdot 9) = 3,55$, что больше критического значения 2,65 и, следовательно, сохраняется сделанный выше вывод о достоверности различия сравниваемых рядов.

§ 4.08. Критерий медианы

В тех случаях, когда неизвестен тип распределения сравниваемых совокупностей или же известно, что оно не является ни нормальным, ни близким к последнему, рекомендуется применение непараметрического критерия медианы. Это целесообразно делать потому, что на положение медианы ряда, в отличие от средней арифметической, величина крайних вариантов, даже резко отклоняющихся, влияет мало.

В качестве примера сравним два сорта земляники по среднему весу плода по 10 растениям сорта x и по 14 растениям сорта y 1) составим общий для двух сортов ранжированный вариационный ряд из данных по среднему весу плода (в граммах) на каждом растении: $y_4, y_8, y_8, x_8, y_9, y_9, x_9, x_{10}, x_{10}, y_{10}, y_{10}, y_{12}, x_{12}, x_{12}, x_{13}, y_{13}, y_{13}, x_{14}, y_{14}, x_{15}, y_{16}, y_{16}, x_{17}, y_{18}$; 2) определим порядковый номер медианной варианты, который равен $0,5(N + 1)$, где N — объем выборки; $0,5(24 + 1) = 12,5$, т. е. в данном примере медиана находится между 12 и 13 членами объединенного ряда (при четном числе вариантов за медиану принимается середина промежутка между двумя центральными вариантами); 3) распределим ряд чисел, приведенных в пункте 1, по клеткам табл. 4.04. Для этого подсчитываем: а) число вариантов x , превышающих медиану по своей величине, б) число вариантов y , превышающих медиану. Затем находим число вариантов x и y , которые располагаются ниже медианы: $c = 8, d = 6$. Подсчитываем суммы, указанные в таблице; 4) дальнейшая обработка данных зависит от величины частот a, b, c, d и от объема выборки. Если хотя

Таблица 4.04. К расчету критерия медианы (медиана = 12,5,
N = 24)

Классы		y	Σ
Выше медианы	$a = 4$	$b = 6$	$4+6=10$
Ниже медианы	$c = 8$	$d = 6$	$8+6=14$
	12	12	24

бы одна из частот будет меньше 5 или равна нулю или объем выборки будет меньше 30, то вычисление ведется по формуле Фишера (4.17), пример расчета по которой также приведен в § 4.14. Если же не окажется частот меньше 5 и будет $N > 30$, то данные можно обработать при помощи обычного критерия хи-квадрат для четырехпольной таблицы (§ 4.14), который несколько проще по расчетам, чем формула (4.17). В нашем примере $N < 30$ и одна из частот $a = 4$, поэтому применяем формулу (4.17), вычисляя через десятичные логарифмы, как это излагается в § 4.14.

$$P(\chi^2) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!}, \quad (4.17)$$

где $P(\chi^2)$ — вероятность того, что в двух сравниваемых выборках из одной генеральной совокупности получится одинаковое распределение частот в четырехпольной таблице; a, b, c, d — частоты, соответствующие ячейкам таблицы 4.04; N — объем выборки ($N = a + b + c + d$); ! — знак факториала. Факториал нуля равен единице: $0! = 1$. $P(\chi^2) = \frac{10!14!12!12!}{24!4!6!8!6!}$; $\lg P(\chi^2) = \bar{1},367$; $P(\chi^2) = 0,23$.

Вероятность того, что распределение частот в табл. 4.04 случайно превышает оба критических значения: 0,01 и 0,05 (см. о процедуре оценки в § 4.14), поэтому нулевую гипотезу считаем справедливой и, следовательно, два сравниваемых сорта земляники по весу плодов достоверно не различаются.

При отсутствии таблиц логарифмов факториалов можно применить формулу Стирлинга (4.18) для приближенного вычисления факториалов:

$$n! = n^{n+0,5} e^{-n} \sqrt{2\pi}, \quad (4.18)$$

где $n!$ — искомый факториал числа n ($1 \cdot 2 \cdot 3 \dots n$); $e = 2,7183$ (основание натуральных логарифмов); $\pi = 3,1416$.

Заменив в формуле (4.17) все факториалы через логарифмы формулы Стирлинга, получим формулу:

$$\begin{aligned} \ln P(\chi^2) = & (a+b+0,5) \ln(a+b) + (c+d+0,5) \ln(c+d) + \\ & (a+c+0,5) \ln(a+c) + (b+d+0,5) \ln(b+d) - \\ & (a+0,5) \ln a - (b+0,5) \ln b - (c+0,5) \ln c - \\ & - (d+0,5) \ln d - (N+0,5) \ln N - \ln \sqrt{2\pi}. \end{aligned} \quad (4.19)$$

Подставив в (4.19) те же, что и выше, числовые значения, получим:

$$\begin{aligned} \ln P(\chi^2) &= (4 + 6 + 0,5) \ln(4 + 6) + (8 + 6 + 0,5) \ln(8 + \\ &+ 6) + (4 + 8 + 0,5) \ln(4 + 8) + (6 + 6 + 0,5) \ln(6 + \\ &+ 6) - (4 + 0,5) \ln 4 - (6 + 0,5) \ln 6 - (8 + 0,5) \ln 8 - \\ &- (6 + 0,5) \ln 6 - (24 + 0,5) \ln 24 - 0,9189 = -1,390; \\ P(\chi^2) &= 0,249, \text{ что подтверждает вывод, сделанный выше.} \end{aligned}$$

Формула (4.19) удобна для программирования на ЭВМ, при помощи которой выполнены приведенные расчеты. Следует отметить, что вычисление через формулу Стирлинга (4.18) дает приближенные результаты, поэтому в сомнительных случаях необходимо проверить выводы, пользуясь формулой Фишера (4.17), логарифмируя ее с помощью таблиц логарифмов факториалов. Однако при возрастании чисел, точность вычисления их факториалов и результатов по формуле (4.19) также возрастает.

§ 4.09. Критерий Вилкоксона

Сравнение совокупностей по их центральной тенденции наиболее просто может быть сделано при помощи непараметрического критерия Вилкоксона, вычисление которого для сравнения по длине колоса двух сортов пшеницы — x и y приводится в табл. 4.05, где по столбцам производятся следующие действия.

Таблица 4.05. К вычислению критерия Вилкоксона

Ранжированные исходные данные по длине колоса у двух сортов, см		Ряды после исключения совпадающих вариантов		Ранг	
x	y	x	y	ряда x	ряда y
1	2	3	4	5	6
	7,4		7,4		1
	7,5		7,5		2
	7,7		7,7		3
7,9*	7,9*		8,0		4,5
	8,0		8,0		4,5
	8,0		8,2		6,5
	8,2		8,2		6,5
	8,2	8,4		8	
8,4		9,0		10	
8,5*	8,5*	9,0		10	
9,0		9,0		10	
9,0		9,3		12	
9,0		9,7		13	
9,3		10,4		14	
9,7		10,5		15	
10,4					
10,5					
				$n_x = 92$	$n_y = 28$
				$N_x = 8$	$N_y = 7$

1 и 2. Совместное ранжирование обоих рядов x и y в возрастающем порядке сверху вниз (можно и снизу вверх). Варианты обоих рядов объединены в один ряд по возрастанию величин, но записаны в разных столбцах во избежание путаницы.

3 и 4. При ранжировании в столбцах 1, 2 оказалось, что варианты 7,9 и 8,5 совпадают в обоих рядах. В таких случаях совпадающие пары вариант исключаются из дальнейших расчетов, что и сделано в столбцах 3, 4.

5 и 6. Объединенному ряду вариант (столбцы 2, 3) присваиваем порядковые номера — ранги. Процедура присваивания рангов описана в § 2.02 пункт 1. В данном случае в столбце 3 имеются три равные варианты: 9,0; 9,0; 9,0, их средний порядковый номер, или ранг, равен: $(9 + 10 + 11)/3 = 10$. В столбце 4 варианты 8,0 и 8,0 получают средний ранг 4,5 каждая, и т. д.

Суммы столбцов 5 и 6: $n_x = 92$ и $n_y = 28$, число членов в рядах: $N_x = 8$; $N_y = 7$, $N = N_x + N_y = 15$. Полученные величины проверим по равенству $n_x + n_y = \frac{N(N+1)}{2}$, $92 + 28 = \frac{15(15+1)}{2}$, $120 = 120$.

Равенство для приведенных величин соблюдается, следовательно, расчеты выполнены правильно.

Обращаясь к табл. 16П, находим, что для $N_x = 8$ и $N_y = 7$ сумма рангов равна 38 при уровне значимости $W'_1 = 0,05$. Для сравнения с таблицей берется всегда меньшая из вычисленных сумм рангов, т. е. в данном случае: $n_y = 28$. Поскольку вычисленное значение n_y меньше табличного: $n = 38$, делаем вывод о достоверности различия между двумя сортами пшеницы по длине колоса. Если вычисленное значение n будет больше табличного, то сравниваемые ряды достоверно не различаются. При $N_x > 10$ или $N_y > 10$, когда нельзя воспользоваться табл. 16П, для оценки различия рядов можно применить критерий, вычисляемый по формуле:

$$t = \frac{N_x(N+1) - 2n_x}{\sqrt{N_x N_y (N+1)}}, \quad (4.20)$$

где t — критерий достоверности различия, его граничные значения при уровнях значимости: $W'_1 = 0,05$ и $W'_2 = 0,01$ соответственно 1,13 и 1,49. Если вычисленное значение критерия превышает эти величины, то различие между рядами достоверно; N_x — число членов того из сравниваемых рядов, который имеет меньшую сумму рангов; N — общее число вариантов в обоих сравниваемых рядах; n_x — меньшая из двух сумм рангов; N_y — число членов того из сравниваемых рядов, который имеет большую сумму рангов.

Для рассмотренного в данном параграфе примера по формуле (4.20) получим:

$$t = \frac{7(15+1) - 2 \cdot 28}{\sqrt{7 \cdot 8(15+1)}} = 1,87.$$

Поскольку вычисленное значение критерия превышает выше-названные его критические значения, делаем заключение о достоверности различия между двумя сравниваемыми сортами пшеницы по длине колоса.

§ 4.10. Критерий ван дер Вардена

При помощи непараметрического критерия ван дер Вардена можно с большей точностью, чем по методу из § 4.09, сравнивать невзвешенные вариационные ряды по их центральной тенденции. По технике вычислений этот критерий, однако, несколько сложнее, чем критерий Вилкоксона (см. § 4.09), и требует применения двух специальных таблиц. Критерий ван дер Вардена вычисляется по формуле

$$X_w = \sum \Psi \left(\frac{x_i}{N+1} \right), \quad (4.21)$$

где X_w — критерий достоверности различия ван дер Вардена; Ψ — функция, обратная интегралу вероятности, ее значения по величине выражения в следующих за ней скобках берутся по табл. 17П; x_i — порядковые номера ранжированных членов того из сравниваемых рядов, который содержит меньше вариантов; N — общая численность вариантов обоих сравниваемых рядов; Σ — знак суммирования.

Рассмотрим технику вычисления критерия ван дер Вардена на примере из § 4.09. В табл. 4.06 по столбцам производятся следующие действия.

Таблица 4.06. К вычислению критерия ван дер Вардена

Длина колоса у двух сортов пшеницы, см		Порядковые номера вариант		Значения выражения: $\frac{x_i}{N+1}$		Значения из табл. 17П $\Psi \left(\frac{x_i}{N+1} \right)$	
x	y	x	y	x	y	x	y
	7,4		1		0,062		—1,54
	7,5		2		0,125		—1,15
	7,7		3		0,188		—0,89
	8,0		4		0,250		—0,67
	8,0		5		0,312		—0,49
	8,2		6		0,375		—0,32
	8,2		7		0,438		—0,16
8,4		8		0,500		0,00	
9,0		9		0,562		0,16	
9,0		10		0,625		0,32	
9,0		11		0,688		0,49	
9,3		12		0,750		0,67	
9,7		13		0,812		0,89	
10,4		14		0,875		1,15	
10,5		15		0,938		1,54	
$N_x=8$ $N_y=7$						5,22	—5,22

1 и 2. Из столбцов 3 и 4 табл. 4.05 переписываем упорядоченные ряды x и y , из которых исключены совпадающие по величине варианты (см. § 4.09 пункт 3.4).

3 и 4. Нумерация порядковыми номерами вариант из столбцов 1, 2.

5 и 6. Вычисление значений выражения $x_i/(N + 1)$ для рядов x и y из столбцов 3, 4. Каждый порядковый номер делим на $N + 1 = 15 + 1 = 16$, например, для ряда y :

$$1/(15 + 1) = 0,062; \quad 2/(15 + 1) = 0,125 \text{ и т. д.}$$

7 и 8. Пользуясь табл. 17П, находим по числам столбцов 5, 6 значения функции $\Psi(x_i/N + 1)$. Суммы чисел обоих столбцов должны совпасть (не принимая во внимание знак), что используется для проверки вычислений.

Полученная сумма 5,22 является вычисленной величиной критерия ван дер Вардена.

По табл. 18П для $N = 15$ и $N_x - N_y = 1$, по 5% уровню значимости находим, что табличное значение критерия ван дер Вардена равно 3,24. Знак критерия во внимание не принимается. Вычисленное значение критерия равно 5,32, следовательно, между сравниваемыми по длине колоса сортами пшеницы имеется достоверное различие.

§ 4.11. Серийный критерий

Серийный критерий, в отличие от критериев, рассмотренных в § 4.09, 4.10, позволяет обнаружить различие между двумя совокупностями не только по центральной тенденции, но и по другим их свойствам. Вычислим серийный критерий для примера из § 4.09.

1. Из табл. 4.05 столбцов 3, 4 переписываем упорядоченные ряды x и y с исключенными вариантами, совпадающими по величине в обоих рядах. Получим следующие ряды:

$$\begin{array}{l} x \qquad \qquad \qquad 8,4 \ 9,0 \ 9,0 \ 9,0 \ 9,3 \ 9,7 \ 10,4 \ 10,5 \\ y \ 7,4 \ 7,5 \ 7,7 \ 8,0 \ 8,0 \ 8,2 \ 8,2 \end{array}$$

2. Подсчитываем число серий в полученном объединенном упорядоченном ряду. Серий (групп) оказалось две, одна из них начинается с 8,4 и кончается 10,5 (ряд x), другая — от 7,4 до 8,2 (ряд y). По табл. 19П при $N_x = 8$ и $N_y = 7$ находим, что критическое число серий в данном случае должно быть 4. Фактическое число серий равно 2. Гипотеза о различии между рядами отвергается, если фактическое число серий будет больше табличного, и различие считается достоверным, если фактическое число серий будет меньше табличного значения без двух. В данном случае фактическое число серий равно 2, т. е. меньше табличного, равного 4, но не меньше табличного значения без двух: $4 - 2 = 2$, поэтому вопрос о различии двух сортов пшеницы по длине колоса, решаемый с помощью серийного критерия, остается открытым.

Если число вариант в одном из сравниваемых рядов больше 20, вследствие чего нельзя будет пользоваться табл. 19П, для оценки достоверности различия между рядами по числу серий можно применить критерий, вычисляемый по формуле

$$t = \frac{(2N_x N_y)/N - n_s + 0,5}{\sqrt{\frac{2N_x N_y (2N_x N_y - N)}{N^2 (N - 1)}}}, \quad (4.22)$$

где t — критерий достоверности различия; N_x, N_y — число вариант в сравниваемых рядах; n_s — число серий в обобщенном упорядоченном ряду $x + y$; $N = N_x + N_y$.

Граничные значения критерия при 5 и 1% уровнях значимости равны соответственно 1,96 и 2,58. Различие между рядами считается достоверным, если вычисленное значение критерия будет больше 2,58, и недостоверным, если оно будет меньше 1,96. Для того же примера, рассмотренного в данном параграфе, критерий достоверности различия равен по формуле (4.22):

$$t = \frac{(2 \cdot 8 \cdot 7)/15 - 2 + 0,5}{\sqrt{\frac{2 \cdot 8 \cdot 7 (2 \cdot 8 \cdot 7 - 15)}{15^2 (15 - 1)}}} = 3,21.$$

Поскольку вычисленное значение критерия больше 2,58, подтверждается сделанный выше вывод о достоверности различия двух сортов пшеницы по длине колоса.

§ 4.12. Критерий знаков

Критерий знаков для сравнения выборок с попарно сопряженными вариантами является самым простым по технике вычислений. Однако он имеет тот недостаток, что им учитывается не величина, а только лишь знак различия попарно связанных вариант. Поэтому критерий знаков рекомендуется применять только тогда, когда величина этих различий не имеет значения или для быстрых предварительных оценок, которые в сомнительных случаях могут быть проверены более точными методами (§ 4.02 или § 4.13).

Схема расчета критерия знаков представлена в табл. 4.07, где по столбцам приведены следующие данные.

1. В столбце 1 выписаны номера междоузлий растений клевера (снизу вверх по стеблю).

2 и 3. Длина междоузлий у двух сортов — А и Б (в см).

4. Из чисел столбца 2 вычитаются соответственно числа столбца 3. Если числа совпадут, то эта пара исключается из обоих рядов и соответственно уменьшается N .

5. Записываются знаки чисел из столбца 4.

Подсчитывается число того из знаков, который представлен в меньшем количестве, в данном случае — число отрицательных знаков, которых в столбце 5 насчитывается $n^- = 2$. По табл. 20П находим, что при объеме выборки $N = 8$ и уровне значимости 5% число зна-

Таблица 4.07. К сравнению выборок при помощи критерия знаков

Номер междоузлия	Средняя длина междоузлия у сортов клевера, см		δ	Знаки δ
	А	Б		
1	2	3	4	5
1	1,0	0,5	0,5	+
2	3,0	1,7	1,3	+
3	9,0	4,2	4,8	+
4	15,0	5,0	10,0	+
5	14,7	8,6	6,1	+
6	9,5	9,4	0,1	+
7	6,7	8,6	-1,9	-
8	3,0	6,1	-3,1	-
$N = 8$		$n^+ = 6; n^- = 2$		

Таблица 4.08. К вычислению парного критерия Вилкоксона

Номер междоузлия снизу вверх	Длина междоузлий (см) у сортов:		Разность $\delta = x - y$	Ранги δ	Сумма разностей рангов со знаком минус
	x	y			
1	2	3	4	5	6
1	1,0	0,5	0,5	2	
2	3,0	1,7	1,3	3	
3	9,0	4,2	4,8	6	
4	15,0	5,0	10,0	8	
5	14,7	8,6	6,1	7	
6	9,5	9,4	0,1	1	
7	6,7	8,6	-1,9	4	4
8	3,0	6,1	-3,1	5	5

$N = 8$

$n_T = 9$

ков должно быть равно 1. Вычисленное число знаков меньшей группы n^- равно 2, следовательно, два сравниваемых сорта по длине междоузлий не различаются. Если бы вычисленное n было меньше табличного, то рассматриваемые сорта могли бы различаться друг от друга по длине междоузлий.

§ 4.13. Парный критерий Вилкоксона

В § 4.09 приводилась методика расчета критерия Вилкоксона для любых выборок. В данном же параграфе рассматривается критерий для применения лишь к попарно сопряженным вариантам. Парный критерий Вилкоксона в отличие от критерия знаков (см. § 4.12), помогает с большей точностью сравнить выборки с попарно сопряженными вариантами, так как учитывает не только знаки разностей вариантов, но и величину этих разностей. Оценка значимости парного критерия Вилкоксона производится по табл. 21П, а при объеме выборок $N > 25$ — с помощью критерия t .

Рассмотрим схему вычисления критерия Вилкоксона для примера из табл. 4.07. В табл. 4.08 производятся следующие действия по столбцам.

1—4. Точно повторяются все действия, указанные в § 4.12.

5. Ранжируем числа столбца 4 без учета знаков, наименьшее число — 0,1 получает ранг 1, а наибольшее — 10,0 получает ранг 8. При совпадающих числах поступаем так, как сказано в § 2.02 (пункт 1), т. е. присваиваем им обоим их средний ранг.

6. Числа в столбце 4 имеют знаки плюс и минус, отбираем ту группу чисел с одинаковым знаком, в которой содержится меньше

чисел, чем в другой группе. Таковой в данном примере является группа со знаком минус, для которой выписываем в столбце 6 ее ранги из столбца 5 и суммируем их. Сумма рангов меньшей группы $n_T = 4 + 5 = 9$. Различие между сравниваемыми рядами считается недостоверным, если вычисленная сумма рангов будет больше табличной, и различие считается достоверным, если вычисленная сумма рангов будет меньше табличной. По табл. 21П при объеме выборки $N = 8$ и 5% уровне значимости сумма рангов равна 4. Поскольку вычисленная сумма рангов $n_T = 9$ больше табличной, делаем вывод об отсутствии достоверного различия между сравниваемыми рядами.

Если объем сравниваемых рядов $N > 25$, то оценка достоверности различия делается по критерию t , так как табл. 21П пользоваться будет нельзя. Критерий t в применении к оценке суммы рангов парного критерия Вилкоксона вычисляется по формуле

$$t = \frac{[N(N+1)]/4 - n_T}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}, \quad (4.23)$$

где t — критерий различия, граничные значения которого 1,96 и 2,58 на уровнях значимости соответственно $W'_1 = 0,05$ и $W'_2 = 0,01$; N — число пар вариант в сравниваемых рядах; n_T — сумма рангов меньшей группы разностей с одинаковым знаком.

Для примера, рассмотренного в данном параграфе:

$$t = \frac{[8(8+1)]/4 - 9}{\sqrt{\frac{8(8+1)(2 \cdot 8 + 1)}{24}}} = 1,26.$$

Поскольку вычисленное значение критерия 1,26 меньше граничного значения 1,96, делаем вывод об отсутствии различия между сравниваемыми рядами. Ряды считаются достоверно различающимися, если вычисленное значение критерия будет больше 2,58. Если величина критерия заключается между граничными значениями, вопрос о достоверности различия рядов остается нерешенным.

§. 4.14. Критерий хи-квадрат

В различных биологических исследованиях при оценке результативности новых методов, ядохимикатов, препаратов, лекарств, сравнении урожайности стандартных и новых сортов, изучении расщепления потомства при скрещивании и многих других случаях научной и производственной деятельности широко применяется критерий хи-квадрат.

В данном параграфе рассмотрено девять вариантов применения этого критерия в биологических исследованиях, которые, однако, не исчерпывают всех случаев использования критерия хи-квадрат в биометрии.

а. Оценка совпадения теоретических и эмпирических частот взвешенных рядов.

Для оценки меры совпадения теоретической кривой распределения с фактическими данными часто применяется в качестве критерия согласия χ^2 -квadrat. Этот критерий может применяться лишь при наличии не менее 5 частот в каждом классе. Если в каком-либо классе вариационного ряда содержится частот меньше, чем 5, то этот класс вместе с частотами следует объединить с соседним классом.

Общая формула вычисления этого критерия:

$$\chi^2 = \frac{(f - f')^2}{f'}, \quad (4.24)$$

где χ^2 — критерий хи-квадрат; f — фактические частоты, или численности; f' — ожидаемые или теоретические частоты, или численности.

В качестве примера найдем меру соответствия фактическим частотам — теоретической кривой биномиального распределения, вычисленной в § 1.27.

В табл. 4.09 по столбцам приводятся следующие данные.

1. Классы распределения проб корнеплодов по поражаемости фомозом.

2. Всего проб одинакового размера было взято 112, по 9 корнеплодов в каждой.

3. Теоретические частоты кривой биномиального распределения вычислены ранее (см. § 1.27). Ввиду того что по условиям применения критерия хи-квадрат требуется, чтобы частота в каждой клетке была не меньше 5, объединяем в столбцах 2 и 3 классы 1, 2 и 3 в один с частотами 11 и 12, а также 8 и 9, в другой с частотами 24 и 16 (по столбцам 2 и 3). Дальше, следовательно, ведем расчеты уже с 6 классами распределения

4. Разность эмпирических и теоретических частот.

Таблица 4.09. К сравнению эмпирических и теоретических частот биномиального распределения критерием хи-квадрат

Число пораженных корнеплодов в пробе, x	Число проб фактическое, f	Число проб теоретическое, f'	$f - f'$	$(f - f')^2$	$\frac{(f - f')^2}{f'}$
1	2	3	4	5	6
0	0	0			
1	0	1			
2	4	3	—1	1	0,083
3	7	8			
4	18	17	1	1	0,059
5	27	24	3	9	0,375
6	14	24	—10	100	4,167
7	18	19	—1	1	0,053
8	20	11	8	64	4,000
9	4	5			
	112	112		$\chi^2 = 8,737$	

5. Квадраты разностей частот.

6. Критерий хи-квадрат по строкам, в итоге столбца 6 получаем искомую величину $\chi^2 = 8,737$.

Для оценки полученной величины хи-квадрат необходимо знать число степеней свободы, которое зависит от того, какой тип теоретического распределения сравнивается с данным фактическим распределением. При нормальном распределении число степеней свободы: $\nu = k - 3$, при биномиальном и пуассоновском $\nu = k - 2$, при распределении Шарлье: $\nu = k - 4$, где k — число классов взвешенного ряда; ν — число степеней свободы.

В табл. 4.09 с эмпирическим сравнивается биномиальное распределение, в котором после укрупнения крайних вариантов стало 6 классов, следовательно: $\nu = 6 - 2 = 4$ степени свободы. По табл. 10П для уровня значимости $W'_1 = 0,05$ и $\nu = 4$ находим табличное значение хи-квадрат, равное 9,488.

Поскольку вычисленное значение, равное 8,737, меньше табличного, делаем вывод, что теоретическое распределение частот несущественно отличается от эмпирического. Поэтому рассматриваемые эмпирические данные в первом приближении могут быть моделированы биномиальным законом распределения, не исключая возможности более близкого соответствия частот другого типа распределения, так как по 0,01 уровню значимости табличное значение хи-квадрат равно 13,277.

Достоверность хи-квадрат можно, особенно при отсутствии таблиц, оценить по правилу Романовского:

$$\frac{\chi^2 - \nu}{\sqrt{2\nu}} \geq 3, \quad (4.25)$$

где ν — число степеней свободы; χ^2 — вычисленное значение хи-квадрат. Различие считается достоверным, если левая часть неравенства будет равна или больше трех. Для рассматриваемого примера: $(8,737 - 4) / \sqrt{2 \cdot 4} = 1,67 < 3$, следовательно, биномиальное распределение подходит к эмпирическим данным.

б. Сравнение дисперсий генеральной совокупности и выборки.

Если известно, что данные изучаемого признака распределены по нормальному закону, то можно посредством критерия хи-квадрат оценивать различие между дисперсией выборки и дисперсией генеральной совокупности по формуле:

$$\chi^2 = \frac{(N - 1) \sigma^2}{\sigma_0^2}, \quad (4.26)$$

где χ^2 — вычисляемая величина критерия; N — объем выборки; σ^2 — дисперсия выборки; σ_0^2 — дисперсия генеральной совокупности.

Исследовалась стабильность комплекса ежегодных агротехнических мероприятий по уходу за одной культурой. По выборке в 2619 растений была вычислена дисперсия значений высоты особей, равная $\sigma^2 = 2009,55$. По данным прошлых лет известно, что величи-

ны измерений высоты особей распределяются по нормальному закону и дисперсия этого признака обычно была близкой к $\sigma_0^2 = 1397,79$. Требуется выяснить достоверно ли различие этих дисперсий. По формуле (4.26)

$$\chi^2 = \frac{(2619 - 1) 2009,55}{1397,79} = 3760.$$

Число степеней свободы в этом примере больше 100, которое приведено в табл. 10Па, поэтому для оценки полученной величины хи-квадрат воспользуемся приближенной формулой:

$$t = \frac{\chi^2 - (N - 1)}{\sqrt{2(N - 1)}}, \quad (4.27)$$

где t — нормированное отклонение; χ^2 — вычисленная величина хи-квадрат; N — объем совокупности, т. е.:

$$t = \frac{3760 - (2619 - 1)}{\sqrt{2(2619 - 1)}} = \frac{1142}{\sqrt{5236}} = \frac{1142}{72,37} = 15,8.$$

По формуле приближения Фишера величина χ^2 оценивается следующим образом:

$$t = \sqrt{2\chi^2} - \sqrt{2N - 3}, \quad (4.28)$$

где обозначения те же.

$$t = \sqrt{2 \cdot 3760} - \sqrt{2 \cdot 2619 - 3} = 86,72 - 72,35 = 14,4.$$

Полученную величину критерия Стьюдента следует оценивать по табл. 3П при числе степеней свободы $\nu = N - 1$. Однако в данном случае вычисленная величина далеко превосходит число 3 и все табличные значения, поэтому делаем вывод о том, что данные измерений высоты растений за последний год резко отличаются, по своей дисперсии от обычной ее величины и, следовательно, комплекс агротехнических условий или условия данного года выращивания сильно отличаются от прошлых лет.

в. Сравнение двух эмпирических взвешенных рядов одинакового объема.

Сравнение двух выборок одинакового объема критерием хи-квадрат следует производить при численности выборки не менее $N \geq 30$, а частоты в классах обоих рядов должны быть не меньше 5. Критерий хи-квадрат при сравнении выборок одинаковой численности вычисляется по формуле:

$$\chi^2 = \sum \frac{(f_x - f_y)^2}{f_x + f_y}, \quad (4.29)$$

где χ^2 — критерий хи-квадрат; f_x, f_y — частоты сравниваемых рядов x и y .

Сравним по срокам зацветания две равночисленные популяции растений. В табл. 4.10 по столбцам¹ приведены следующие данные.

Таблица 4.10. Сравнение критерием хи-квадрат двух равночисленных выборок

Дата наблюдения	Число зацветших растений в популяции x , f_x	Число зацветших растений в популяции y , f_y	$f_x - f_y$	$f_x + f_y$	$\frac{(f_x - f_y)^2}{f_x + f_y}$
1	2	3	4	5	6
1.VI	85	96	—11	181	0,669
6.VI	156	169	—13	325	0,520
11.VI	132	134	—2	266	0,015
16.VI	81	58	23	139	3,806
21.VI	31	28	3	59	0,153
	485	485	0	970	5,163

1. Сроки наблюдений через 5 дней.

2 и 3. Число зацветших растений в популяции x по пятидневкам и то же самое в популяции y .

4. Попарные разности чисел столбцов 2 и 3.

5. Попарные суммы чисел столбцов 2 и 3.

6. Квадраты чисел столбца 4, деленные на числа столбца 5 по тем же строкам. Сумма чисел столбца 6 есть результат вычислений по формуле (4.29). Число степеней свободы в подобных случаях равно: $\nu = k - 1 = 5 - 1 = 4$, т. е. числу классов без одного. По табл. 10П при $\nu = 4$ и на уровне значимости 0,05 табличное значение хи-квадрат равно 9,488, а вычисленная его величина, составляющая 5,163, меньше табличной; следовательно, две сравниваемые популяции по срокам зацветания не различаются.

г. Сравнение двух эмпирических взвешенных рядов разного объема.

При сравнении двух эмпирических распределений критерием хи-квадрат применяется несколько иная схема его расчета, так как неизвестен тип этих распределений. При разном объеме рядов эмпирических распределений для вычисления хи-квадрат как критерия различия между ними употребляется формула:

$$\chi^2 = \frac{\sum \frac{(f_x N_y - f_y N_x)^2}{f_x + f_y}}{N_x N_y}, \quad (4.30)$$

где χ^2 — критерий различия хи-квадрат; f_x, f_y — частоты рядов x и y ; N_x, N_y — объемы сравниваемых рядов x и y .

Пример вычисления хи-квадрат по формуле (4.30) приводится в табл. 4.11, где по столбцам приведены следующие данные.

1. Число археспориальных клеток, которое колебалось от 1 до 6 в семяпочках растений томата.

2 и 3. Число семяпочек с тем или иным числом археспориальных клеток в них у сорта томата x и соответственно у сорта y . Так как по условиям применения критерия хи-квадрат отдельные частоты не

Таблица 4.11. К сравнению критерием хи-квадрат двух эмпирических взвешенных рядов разного объема

Число клеток	Число семян (частоты) у сор- тов		$f_x N_y$	$f_y N_x$	$f_x N_y -$ $- f_y N_x$	$(f_x N_y - f_y N_x)^2$	$\frac{(f_x N_y - f_y N_x)^2}{f_x + f_y}$
	f_x	f_y					
1	2	3	4	5	6	7	8
1	336	252	125 328	136 836	11 508	132 434 064	225 228,00
2	148	102	55 204	55 386	182	33 124	132,50
3	31	12	11 563	6 516	5 047	25 472 209	592 376,95
4	24	5	10 444	3 801	6 643	44 129 449	1 260 841,40
5 } 4	3 } 28	2 } 7					
6 } 1	1	0					

$$N_x = 543 \quad N_y = 373$$

$$\Sigma = 2078578,85$$

должны быть меньше 5, объединяем классы 4, 5 и 6 в один четвертый, в результате чего объединяются частоты $24 + 3 + 1 = 28$ в столбце 2 и частоты $5 + 2 = 7$ в столбце 3. Сумма частот в этих столбцах $N_x = 543$, $N_y = 373$.

4. Перемножаем на $N_y = 373$ поочередно все четыре числа столбца 2.

5. Перемножаем на $N_x = 543$ поочередно все 4 числа столбца 3.

6. Находим попарные разности чисел столбцов 4 и 5.

7. Возводим в квадрат числа столбца 6.

8. Числа столбца 7 делим на сумму частот в данной строке и выписываем результаты в столбце 8, например: $33124/(148 + 102) = 132,50$.

Сумму столбца 8, равную 2078578,85, делим на произведение объемов сравниваемых совокупностей по формуле (4.30):

$$\chi^2 = \frac{2078578,85}{373 \cdot 543} = 10,26.$$

Число степеней свободы при сравнении двух эмпирических распределений равно $\nu = k - 1$, где ν — число степеней свободы; k — число классов распределения; в данном случае: $\nu = 4 - 1 = 3$.

По табл. 10П при $W_1' = 0,05$ и $W_2' = 0,01$ табличное значение хи-квадрат соответственно: 7,815; 11,345. Вычисленное значение хи-квадрат больше первого уровня значимости, но меньше второго, поэтому различие двух сортов томата по распределению археспоридных клеток нельзя считать достоверным. Если бы вычисленное значение хи-квадрат было больше также и 11,345 (по второму уровню значимости), то тогда можно было бы с большей определенностью считать достоверно различающимися два сравниваемых ряда.

д. Применение критерия хи-квадрат при изучении расщепления гибридов.

В генетических и других биологических исследованиях часто возникает необходимость оценки соответствия распределения осо-

бей определенному численному соотношению. Например, требуется установить, произошло ли расщепление потомства гибрида по тому или иному признаку в соответствии с законами Менделя, в других случаях надо оценить достоверность влияния какого-либо фактора на появление мутантов в последующих генерациях.

Для оценки совпадения ожидаемых и фактических численностей особей с определенными признаками в подобных случаях можно применить критерий хи-квадрат.

Рассмотрим применение этого критерия для оценки фактического соотношения численностей особей в потомстве по наблюдаемому признаку. В потомстве изучаемого растения оказалось $f_x = 250$ мужских особей и $f_y = 206$ женских, общая численность потомства $N = 456$. Требуется проверить, соответствуют ли эти численности теоретически ожидаемому соотношению 1 : 1. Вычисления ведутся в следующем порядке.

1. Определим теоретическую численность для бóльшей фактической численности, т. е. долю мужских особей, которая равна:

$$f'_x = \frac{1}{2} N; \quad f'_x = \frac{1}{2} \cdot 456 = 228.$$

2. Теоретическая доля меньшей численности:

$$f'_y = \frac{1}{2} N; \quad f'_y = \frac{1}{2} \cdot 456 = 228.$$

3. По формуле (4.24) определяем хи-квадрат каждой численности и его сумму:

$$\chi^2 = \frac{(250 - 228)^2}{228} + \frac{(206 - 228)^2}{228} = 4,246.$$

4. Число степеней свободы в подобных случаях (при $k = 2$, т. е. двух классах или группах распределения) равно: $\nu = k - 1 = 2 - 1 = 1$. По табл. 10П при $\nu = 1$ для 0,05 уровня значимости находим табличное хи-квадрат, равное 3,841, что меньше вычисленного (4,246); следовательно, в потомстве изучаемого растения имеется значимое отклонение фактического соотношения мужских и женских особей от ожидаемого теоретического соотношения 1 : 1.

В подобных случаях можно применить также и более простую формулу:

$$\chi^2 = \frac{(f_x - df_y)^2}{d(f_x + f_y)}, \quad (4.31)$$

где χ^2 — критерий различия хи-квадрат; f_x — фактическая численность бóльшей группы; f_y — фактическая численность меньшей группы; d — ожидаемое теоретическое соотношение численностей: $d = f'_x/f'_y$, т. е. отношение бóльшей теоретической численности f'_x к меньшей теоретической численности f'_y .

Исследователь далее предположил, что полученные им фактические численности мужских и женских цветков в потомстве ближе соот-

ветствуют теоретическому соотношению 5 : 4 (мужские цветки : женские цветки), откуда $d = 5/4 = 1,25$.

По формуле (4.31) получим:

$$\chi^2 = \frac{(250 - 1,25 \cdot 206)^2}{1,25 (250 + 206)} = 0,099.$$

Проверим этот результат по формуле (4.24). Доля мужских цветков:

$$\frac{5}{9} \cdot 456 = 253,33.$$

Доля женских цветков:

$$\frac{4}{9} \cdot 456 = 202,67.$$

$$\chi^2 = \frac{(250 - 253,33)^2}{253,33} + \frac{(206 - 202,67)^2}{202,67} = 0,044 + 0,055 = 0,099.$$

Результаты совпали. Таким образом, в потомстве изучаемого растения наблюдается большая тенденция к соотношению мужских и женских цветков не 1 : 1, а к 5 : 4, так как в последнем случае величина хи-квадрат меньше ($4,246 > 0,099$).

Критерий хи-квадрат можно применять во многих случаях проверки соответствия соотношения фактических численностей ожидаемым. Часто он используется в психологических тестах.

Например, исследовали вопрос о соотношении числа четных и нечетных цифр в сотне наугад названных цифр (от 0 до 9) у 5 лиц. Соотношения числа четных цифр к числу нечетных у них оказались следующими: 41 : 59; 61 : 39; 50 : 50; 52 : 48; 52 : 48. В той же последовательности эти лица чаще называли следующие цифры: (1), (0), (5 и 8), (4), (0 и 1). Реже, чем остальные, назывались цифры: (8), (1), (1 и 4), (0), (7 и 8). По формуле (4.31), где в данном случае принимаем $d = 50/50 = 1$, в том же порядке получены значения хи-квадрата: 3,24; 4,84; 0; 0,16; 0,16. Для 0,05 уровня значимости при $v = 1$ в табл. 10П хи-квадрат равен 3,841. Сравнивая полученные значения с табличным, видим, что второй субъект ($\chi^2 = 4,84$) отдавал предпочтение четным числам (61 число из 100). У остальных лиц отклонения от теоретического соотношения были несущественными.

По тому же принципу, как это сделано выше при отношении численностей 1 : 1, оценивается по хи-квадрат соотношение фактических и теоретических численностей при любых других заданных теоретических соотношениях: 9 : 3 : 3 : 1 или 12 : 3 : 1 и т. д. При этом сначала вычисляются дроби для теоретических численностей, например знаменатель дроби: $9 + 3 + 3 + 1 = 16$, а дроби для умножения на объем выборки будут следующими: $9/16N$; $3/16N$; $3/16N$; $1/16N$; далее расчеты ведутся по формуле (4.24).

е. Применение критерия хи-квадрат в случае четырехпольной таблицы вариантов.

Под четырехпольной таблицей вариант понимается таблица $2 \times 2 = 4$ классов распределения по типу табл. 4.12. Большинство

Таблица 4.12. Зависимость величины тела бабочек от жизненной формы растений, на которых они обитают

Величина тела бабочек	Жизненная форма растений		Σ
	Древесные	Травянистые	
Крупные	$a = 84$	$b = 53$	$a + b = 137$
Мелкие	$c = 24$	$d = 46$	$c + d = 70$
	$a + c = 108$	$b + d = 99$	$N = 207$

корреляционных зависимостей с многоклассовым распределением в рядах может быть сведено к четырехпольным таблицам за счет укрупнения классов. Вычисление показателей связи при корреляционной решетке 2×2 упрощается, однако применять классы распределения 2×2 рекомендуется лишь тогда, когда не представляет интереса форма изучаемой зависимости.

Снижение числа классов и других группировок в обрабатываемых рядах, как правило, ведет к уменьшению достоверности выводов и потере иногда ценных подробностей. Поэтому без особой необходимости не следует группировать данные в четырехпольную таблицу, если можно составить взвешенный ряд.

Применение хи-квадрат для оценки достоверности коэффициента сопряженности признаков в таблице 2×2 было описано в § 2.13. Здесь без вычисления коэффициента сопряженности будут рассмотрены различные случаи применения хи-квадрат к таблице 2×2 , встречающиеся в биологических исследованиях. Если при изучении корреляционной связи между признаками важно выяснить только степень ее достоверности, то последнюю можно определить по критерию хи-квадрат с поправкой Йейтса:

$$\chi^2 = \frac{(|ad - bc| - 0,5N)^2 N}{(a + b)(c + d)(a + c)(b + d)}, \quad (4.32)$$

где $|ad - bc|$ — означает, что выражение, заключенное в прямые скобки, принимается всегда со знаком плюс, независимо от того, какой знак получается при вычитании внутри скобок; a, b, c, d — частоты в соответствующих ячейках табл. 4.12; N — объем выборки; χ^2 — критерий хи-квадрат.

Число степеней свободы в таблицах 2×2 равно единице, поэтому для оценки величины хи-квадрат требуется только одна верхняя строка из табл. 10П.

Применив критерий хи-квадрат к данным табл. 4.12, получим:

$$\chi^2 = \frac{(|84 \cdot 46 - 53 \cdot 24| - 0,5 \cdot 207)^2 \cdot 207}{(84 + 53)(24 + 46)(84 + 24)(53 + 46)} = 12,5.$$

При $v = 1$ и на любом стандартном уровне значимости (табл. 10П) вычисленный критерий хи-квадрат достоверен, поэтому можно ут-

Таблица 4.13. Частоты появления мутантов у ячменя

Вариант опыта	Число мутантов	Число обычных растений	Σ
Контроль	$a = 0$	$b = 176$	176
Семена подвергались специальной обработке	$c = 5$	$d = 173$	178
	5	349	$N = 354$

верждать, что более крупные бабочки чаще обитают на более крупных растениях.

В тех случаях, когда хотя бы в одной из ячеек комбинационной таблицы частота меньше чем 5, или $N < 30$, необходимо применить для оценки достоверности различия выборок формулу (4.17).

Проверим гипотезу о случайности появления мутантов по данным табл. 4.13, применяя формулу (4.17):

$$P(\chi^2) = \frac{(0 + 176)! (5 + 173)! (0 + 5)! (176 + 173)!}{354! 0! 176! 5! 173!}.$$

После сокращения 5! и 176! получаем:

$$P(\chi^2) = \frac{178! 349!}{354! 173!}.$$

По таблицам [например, (Митропольский, 1952, табл. 24)] находим логарифмы факториалов этих чисел и складываем логарифмы с учетом знака:

$$\lg 178! = 324,7948459$$

$$\lg 349! = 737,5479062$$

$$-\lg 354! = 750,2806020$$

$$-\lg 173! = 313,5673527$$

$$\lg P(\chi^2) = -1,5052026;$$

$$P(\chi^2) = 0,031223.$$

Оценка этой вероятности производится по двум граничным значениям. Если она будет меньше 0,01, то нулевая гипотеза отвергается и считается достоверным различие между сравниваемыми выборками. Если $P(\chi^2)$ будет больше 0,05, то нулевая гипотеза принимается и считается, что сравниваемые выборки не различаются. В нашем случае вычисленная величина 0,031 меньше, чем 0,05, но больше, чем 0,01, поэтому различие между выборками из табл. 4.13 достоверно лишь на 5% уровне значимости, а при более точном подходе следует считать, что вопрос о различии выборок остается открытым.

В биологических исследованиях может встретиться необходимость сравнить выборки с частотами больше 1000, когда нельзя применить

Таблица 4.14. Влияние условий выращивания на появление мутантов у ячменя

Вариант выращи- вания	Число нормаль- ных растений	Число мутантов	Сумма
Контроль	$a = 297$	$b = 0$	297
На агрофоне с бором	$c = 984$	$d = 66$	1050
	1281	66	$N = 1347$

некоторые таблицы логарифмов факториалов. В таких случаях для вычисления факториала можно воспользоваться формулой Стирлинга (4.18). Чем больше n , тем точнее результаты, получаемые по формуле (4.18).

Оценим достоверность появления многоузлых многолистных мутантов ячменя в последующих поколениях после специальной обработки семян и при выращивании на агрофоне с бором (табл. 4.14).

Применяя формулу (4.17), получим:

$$P(\chi^2) = \frac{297! 1050! 1281! 66!}{1347! 297! 0! 984! 66!}.$$

Сокращая $66!$ и $297!$ и заменяя факториал каждого числа выражением по формуле (4.18), то же самое получим в следующем виде:

$$P(\chi^2) = \frac{1050^{1050} e^{-1050} (2\pi)^{0,5} 1050^{0,5} 1281^{1281} e^{-1281} (2\pi)^{0,5} 1281^{0,5}}{1347^{1347} e^{-1347} (2\pi)^{0,5} 1347^{0,5} 984^{984} e^{-984} (2\pi)^{0,5} 984^{0,5}}.$$

Сделав сокращения, логарифмируем это выражение:

$1050 \lg 1050 = 3172,2488$	$1347 \lg e = 584,9886$
$-1050 \lg e = -456,0045$	$-0,5 \lg 1347 = -1,5647$
$0,5 \lg 1050 = 1,5106$	$-984 \lg 984 = -2945,1072$
$1281 \lg 1281 = 3980,7704$	$984 \lg e = 427,3414$
$-1281 \lg e = -556,3255$	$-0,5 \lg 984 = -1,4965$
$0,5 \lg 1281 = 1,5538$	$\lg P(\chi^2) = -7,3430$
$-1347 \lg 1347 = -4215,2582$	$P(\chi^2) = 45 \cdot 10^{-8}$

Поскольку вычисленное значение вероятности значительно меньше $0,01$, делаем вывод о достоверном различии выборок. Следовательно, специальная обработка совместно с добавкой бора к почве влияет на появление мутантов у ячменя.

В заключение отметим, что формулы (4.32), (4.17) вполне могут заменить методы сравнения долей, описанные в § 4.04, 4.05.

ж. Сравнение нескольких совокупностей данных.

Сравнение двух и более выборок одновременно применяется для проверки степени их однородности или принадлежности к одной или разным генеральным совокупностям. Например, перед получением

Таблица 4.15. Сравнение нескольких выборок разного объема критерием хи-квадрат

Сроки наблюдений	Число цветущих растений в географических пунктах			Сумма
	f_x	f_y	f_z	
30. VII	98 (9604) 91	85 (7225) 94	96 (9216) 94	279
10. VIII	166 (27556) 160	156 (24336) 165	169 (28561) 166	491
20. VIII	126 (15876) 127	132 (17424) 132	134 (17956) 133	392
30. VIII	60 (3600) 65	81 (6561) 67	58 (3364) 67	199
20. IX	17 (289) 24	28 (784) 24	28 (784) 25	73
	$N_x = 467$	$N_y = 482$	$N_z = 485$	$N = 1434$

средней арифметической урожая по делянкам стандартного сорта важно определить, можно ли объединять такие данные для приведения к среднему стандарту. При объединении массивов фенологических наблюдений также важно установить степень их однородности. Вместо того, чтобы сравнивать такие выборки попарно, можно сравнить их все одновременно. Условием применения критерия хи-квадрат является, как и прежде, достаточная величина частот: не меньше 5 в каждой клетке таблицы. В табл. 4.15 приведено вычисление хи-квадрат при определении однородности трех выборок. Порядок действий при этом следующий.

1. Возводим в квадрат эмпирическую частоту в каждой клетке таблицы и результат записываем в той же клетке ниже эмпирической частоты в скобках.

2. Находим теоретические частоты в каждой клетке таблицы. Для этого перемножаем итоги строки и столбца, в которых находится данная эмпирическая частота и делим произведение на $N = 1434$. Например:

$$91 \approx \frac{279 \cdot 467}{1434}; \quad 94 \approx \frac{279 \cdot 482}{1434}; \quad 94 \approx \frac{279 \cdot 485}{1434}; \quad 160 \approx \frac{491 \cdot 467}{1434}$$

и т. д.

Теоретические частоты записываем в тех же клетках, ниже чисел в скобках.

3. Величина критерия хи-квадрат определяется по формуле (1.177):

$$\chi^2 = \sum \frac{f^2}{f'} - N,$$

где χ^2 — критерий хи-квадрат; f — эмпирические частоты всех сравниваемых выборок; f' — теоретические частоты; N — объем всех сравниваемых выборок: $N = N_x + N_y + N_z$; N_x, N_y, N_z — объемы трех сравниваемых рядов в табл. 4.15.

По формуле (1.177):

98^2	$91 = 105,54$	60^2	$65 = 55,38$
85^2	$94 = 76,86$	81^2	$67 = 97,93$
96^2	$94 = 98,04$	58^2	$67 = 50,21$
166^2	$160 = 172,23$	17^2	$24 = 12,04$
156^2	$165 = 147,49$	28^2	$24 = 32,67$
169^2	$166 = 172,05$	28^2	$25 = 31,36$
126^2	$127 = 125,01$	$\Sigma f^2/f' = 1443,82;$	
132^2	$132 = 132,00$	$\chi^2 = 1443,82 - 1434,00 = 9,82.$	
134^2	$133 = 135,01$		

4. Число степеней свободы в подобных случаях равно числу сравниваемых выборок без единицы, помноженному на число классов без единицы:

$$v = (3-1)(5-1) = 8.$$

На уровне значимости 0,05 табличная величина хи-квадрат равна 15,507 (табл. 10П), а вычисленное фактическое значение составляет 9,82, т. е. меньше табличного, поэтому делаем вывод об отсутствии различия между сравниваемыми тремя рядами из табл. 4.15. Следовательно, три географических пункта — x, y, z — не различаются по срокам зацветания в них изучаемого вида растения.

3. Оценка критерия согласия без таблиц.

При оценке соответствия частот теоретического и эмпирического рядов распределения при помощи величины хи-квадрат последнюю называют часто критерием согласия Пирсона. При массовой обработке числовых материалов на ЭВМ или в других случаях иногда удобно избежать обращения в процессе счета к таблицам для оценки критерия согласия хи-квадрат. Величину критерия согласия можно оценить по вышеприведенной формуле (4.25), полученной В. И. Романовским, а также по формуле Ястремского:

$$\frac{\chi^2 - k}{\sqrt{2k - 4\theta}} \geq 3, \quad (4.33)$$

где χ^2 — величина хи-квадрат по формуле (4.24); k — число классов распределения; θ — поправка на число классов, при $k \leq 20$ она равна 0,6.

Для величины хи-квадрат, равной 8,737, при числе классов $k = 5$, по формуле (4.33):

$$\frac{8,737 - 5}{\sqrt{2 \cdot 5 + 4 \cdot 0,6}} = 1,06 < 3.$$

Так как вычисленная величина критерия согласия Ястремского меньше 3, делаем вывод о соответствии теоретического распределения эмпирическому.

и. Критерий гомогенности Брандта—Снедекора.

Этот критерий применяется для проверки гипотезы о том, что две выборки данных в виде взвешенных вариационных рядов можно отнести к одной общей совокупности. Величина хи-квадрат определяется в данном случае по формуле:

$$\chi^2 = \frac{N_{xy}^2}{N_x N_y} \left(\sum_1^k \frac{f_x^2}{f_{xy}} - \frac{N_x^2}{N_{xy}} \right), \quad (4.34)$$

где N_x , N_y — сумма частот сравниваемых рядов; N_{xy} — сумма частот обоих рядов, $N_{xy} = N_x + N_y$; f_x — частоты первого ряда; f_{xy} — частоты объединенного ряда; k — число классов распределения.

Алгоритм вычисления критерия гомогенности покажем на примере данных из табл. 4.11, где $N_x = 543$, $N_y = 373$, $N_{xy} = 543 + 373 = 916$. Частоты объединенного ряда равны построчным суммам частот обоих рядов: $336 + 252 = 588$, $148 + 102 = 250$, $31 + 12 = 43$, $24 + 5 = 29$, $4 + 2 = 6$. По формуле (4.34):

$$\chi^2 = \frac{916^2}{543 \cdot 373} \left(\frac{336^2}{588} + \frac{148^2}{250} + \frac{31^2}{43} + \frac{24^2}{29} + \frac{4^2}{6} - \frac{543^2}{916} \right) = 10,8.$$

При числе степеней свободы: $\nu = k - 1 = 5 - 1 = 4$ по табл. 10П критическое значение для уровня значимости 0,05 равно 9,488, что подтверждает вывод, полученный в § 4.14, г.

§ 4.15. Сравнение частот взвешенных рядов по критерию Колмогорова

По своей теоретической основе этот критерий близок к изложенному в § 4.07 критерию лямбда, который иначе называют критерием Колмогорова—Смирнова. При помощи критерия Колмогорова сравнивают лишь взвешенные ряды, как теоретические с эмпирическими, так и эмпирические между собой. Число наблюдений при этом должно быть достаточно большим, объединять их в очень крупные разряды (классы) нельзя, а сама изучаемая случайная величина должна иметь распределение непрерывного типа. В результате вычислений получают величину $K(\lambda)$, которую сравнивают в зависимости от принятого доверительного уровня с одним из трех следующих чисел: 1,36 ($P'_1 = 0,95$); 1,63 ($P'_2 = 0,99$); 1,95 ($P'_3 = 0,999$). Если вычисленная величина критерия будет меньше критического уровня для избранной доверительной вероятности (указана в скобках), то различие считается несущественным, если больше — то устанавливается факт неслучайного различия между двумя данными рядами. Критерий Колмогорова более удобен, чем критерий хи-квадрат, в частности тем, что позволяет обходиться без таблиц, учета числа степеней свободы, объединения классов распределения с малыми частотами,

Таблица 4.16. Расчет критерия Колмогорова при сравнении эмпирических и теоретических частот

Частоты				$\frac{n}{N}$	$\frac{n'}{N}$	d
эмпирические	теоретические	накопленные эмпирические, n	накопленные теоретические, n'			
1	2	3	4	5	6	7
2	2,98	2	2,98	0,00935	0,013925	0,0045
30	17,4	32	20,38	0,14953	0,095234	0,0543
34	49,7	66	70,08	0,30841	0,32748	0,01907
62	71,3	128	141,38	0,59813	0,66065	0,0625
74	50,9	202	192,28	0,94393	0,89850	0,0454
8	18,1	210	210,38	0,98131	0,98308	0,0018
4	3,2	214	213,58	1,000	0,99804	0,002

$N = 214 \quad 213,58$

1. Приведем пример вычисления критерия Колмогорова по данным из § 1.25 (табл. 4.16), где сравниваются частоты эмпирического ряда распределения продолжительности вегетационного периода у сортов ячменя (столбец 1) с частотами соответствующей нормальной кривой (столбец 2). Частоты располагаются в таблице, сверху вниз в порядке возрастания вариантов ряда, что принято в качестве условия при вычислении этого критерия. В столбцах 3 и 4 приведены эти же частоты, но нарастающим итогом, т. е. в столбце 3: $2 + 30 = 32$; $32 + 34 = 66$; $66 + 62 = 128$ и т. д., также и в столбце 4; суммирование частот ведется только сверху вниз. Последнее число в столбце 3 должно быть равно объему эмпирического ряда $N = 214$, а последнее число в столбце 4 должно быть равно сумме частот теоретического ряда $N' = 213,58$. В столбцах 5 и 6 те и другие частоты поделены на $N = 214$, т. е. на объем эмпирического ряда. В столбце 7 вычислена разница без знака построчно между аналогичными частотами эмпирического и теоретического рядов. Для дальнейших расчетов необходима лишь одна разность из этого столбца — максимальная, которая в данном случае равна 0,0625. По формуле (1.178):

$$K(\lambda) = d_{\max} \sqrt{N},$$

где $K(\lambda)$ — величина критерия Колмогорова; d_{\max} — максимальная разница по модулю числа (без знака) из столбца 7; N — объем эмпирического ряда; $K(\lambda) = 0,0625 \sqrt{214} = 0,915$. Это число меньше, чем 1,36 при $P_1 = 0,95$, поэтому делаем вывод, что совпадение частот эмпирического ряда с частотами нормальной кривой удовлетворительное.

2. Посредством критерия Колмогорова можно сравнить также и другие любые два ряда с одинаковыми вариантами, в том числе два эмпирических. Рассмотрим технику вычисления критерия в этом

Таблица 4.17. Сравнение двух эмпирических рядов по критерию Колмогорова

Границы класса (дни от 1 марта)	Частоты (число плодо- носящих видов)		Накопленные частоты		Частоты		Разности 6-7
	1958 г. f_x	1960 г. f_y	n_x	n_y	$\frac{n_x}{N_x}$	$\frac{n_y}{N_y}$	
1	2	3	4	5	6	7	8
88,5—105,4	0	4	0	4	0	0,0214	0,0214
105,5—122,4	5	6	5	10	0,0266	0,0535	0,0269
122,5—139,4	6	4	11	14	0,0585	0,0749	0,0164
139,5—156,4	5	20	16	34	0,0851	0,1818	0,0967
156,5—173,4	25	24	41	58	0,2181	0,3102	0,0921
173,5—190,4	30	25	71	83	0,3776	0,4438	0,0662
190,5—207,4	28	43	99	126	0,5266	0,6738	0,1472
207,5—224,4	55	51	154	177	0,8191	0,9465	0,1274
224,5—241,4	20	10	174	187	0,9255	1,0	0,0745
241,5—258,4	13	0	187	187	0,9947	1,0	0,0053
258,5—275,5	1	0	188	187	1,0	1,0	0,0

$$N_x = 188 \quad N_y = 187$$

случае на примере. При анализе фенологических наблюдений, проведенных в дендрарии Ботанического института АН СССР в Ленинграде, было предположено, что созревание плодов у видов деревьев и кустарников в 1958 г. наступило в целом позже, чем в 1960 г. Проверка этой гипотезы по критерию Колмогорова проведена следующим образом.

Частоты в табл. 4.17, так же как и в предыдущем алгоритме, располагаются сверху вниз от меньшего класса к большему, их суммирование тоже ведется при этом сверху вниз. Однако частоты в столбцах 6 и 7 получают путем деления накопленных частот на объемы только своих рядов: в столбце 6 на N_x , а в столбце 7 на N_y . Полученная в столбце 8 максимальная разность без знака подставляется в формулу:

$$K(\lambda) = d_{\max} \sqrt{\frac{N_x N_y}{N_x + N_y}}, \quad (4.35)$$

где $K(\lambda)$ — критерий Колмогорова; d_{\max} — максимальная разность между частотами обоих рядов; N_x, N_y — объемы сравниваемых эмпирических рядов.

$$K(\lambda) = 0,1472 \sqrt{\frac{188 \cdot 187}{188 + 187}} = 1,42.$$

Полученное число больше критического числа на доверительном уровне $P_1 = 95\%$; $1,42 > 1,36$, но меньше числа 1,63, соответствующего следующему доверительному уровню $P'_2 = 0,99$, поэтому различие по наступлению времени плодоношения растений между 1958

и 1960 гг. считаем не строго доказанным, хотя и можно утверждать, что оно существует с риском допустить ошибку в 5 случаях из 100. Привлечение дополнительных наблюдений может в данном случае повысить надежность вывода.

В § 1.36 дается несколько иной способ оценки достоверности критерия Колмогорова по табл. 28П вероятностей $1 - K(\lambda)$.

§ 4.16. Двухфакторный дисперсионный анализ с повторностями

Результаты полевых опытов, например данные по урожайности, никогда в точности не соответствуют степени влияния на них какого-либо организованного в данном опыте фактора: качества или количества удобрения, сортовых, экологических и других различий. Происходит это потому, что на результаты опыта всегда влияют не только эти, но и другие неучтенные факторы. Задача отделения доли влияния на результаты опыта этих неучтенных факторов от доли влияния интересующих исследователя факторов была решена Р. А. Фишером путем дисперсионного анализа, разработанного впервые для математической обработки данных полевых опытов. В дальнейшем дисперсионный анализ стали успешно применять также во многих других областях науки и практической деятельности. В результате дисперсионного анализа могут быть получены следующие данные: степень достоверности действия на объект каждого из учтенных факторов в отдельности и в их сочетаниях, доля влияния на объект каждого фактора в отдельности, в их сочетаниях и доля влияния неучтенных, случайных факторов, ошибка опыта, ошибка разности вариантов и повторностей опыта, показатель точности опыта.

Как показывает практика, основную трудность для начинающих представляет составление комбинационной таблицы, определение, назначение, разграничение и размещение в таблице градаций факторов и повторностей, т. е. сама организация схемы дисперсионного анализа. Поэтому для уяснения принципов дисперсионного анализа целесообразно ознакомиться сначала с простейшей схемой двухфакторного дисперсионного анализа, в которой тем не менее участвуют все обычные его элементы и где порядок расчетов прослеживается более четко из-за меньшего количества вычислений.

* * *

При включении в дисперсионный анализ двух и большего числа изучаемых факторов следует иметь в виду, что они должны быть независимыми друг от друга, т. е. между ними должна отсутствовать корреляционная связь. Ниже приводятся расчеты по дисперсионному анализу данных, сведенных в четырехпольную табл. 4.18. При этом на урожайность двух сортов томата учтено влияние двух факторов: особенности года выращивания (фактор 1) и особенности сортов (фактор 2). Поэтому проводимый здесь дисперсионный анализ является двухфакторным. Каждый из этих двух факторов имеет две

Таблица 4.18. Дисперсионный анализ влияния метеорологических условий года выращивания (фактор 1) и особенностей сортов томата (фактор 2) на их урожайность (комбинационная таблица 2×2)

Сорт	Урожай, кг/куст		S_2	$\sum_2 x^2$	S_2^2	M_2
	1964 г.	1965 г.				
I	2 3 2 <hr/> 7	2 2 3 <hr/> 7	14	34	196	2,33
II	3 3 2 <hr/> 8	3 4 2 <hr/> 9	17	51	289	2,83
S_1	15	16	$\sum x = 31$	$\sum S_2^2 = 485$		
$\sum_1 x^2$	39	46	$\sum x^2 = 85$			
S_1^2	225	256	$\sum S_1^2 = 481$			
M_1	2,50	2,67				

градации (что является их минимальным числом), вследствие чего таблица данных называется четырехпольной, или таблицей 2×2 . По каждому сочетанию факторов имеются три наблюдения (минимальное число — два наблюдения), т. е. по каждому сорту за каждый отдельный год имеются три повторности, поэтому излагаемый здесь дисперсионный анализ более полно можно называть двухфакторным с тремя повторностями для комбинационной таблицы 2×2 . Эта схема расчета может быть также применена для двухфакторного дисперсионного анализа и с большим числом градаций факторов и повторностей.

Порядок действий по каждому виду дисперсионного анализа определяется его основной задачей: разложить суммарное или общее варьирование изучаемого объекта на доли, которые в нем составляют: варьирование, вызываемое действием отдельных факторов, варьирование, вызываемое прямым взаимодействием факторов между собой, и остаточное варьирование объекта, зависящее от неучитываемых факторов. Измеряя колебания вариант данной выборки суммой квадратов отклонений от средних арифметических по каждому фактору, задачу двухфакторного дисперсионного анализа можно представить в виде равенства, слагаемые которого требуется найти:

$$\theta = \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5, \quad (4.36)$$

где θ — общее суммарное варьирование; θ_1 — варьирование, вызываемое фактором 1; θ_2 — варьирование, вызываемое фактором 2; θ_3 — варьирование, возникающее только вследствие взаимодействия

факторов 1 и 2; θ_4 — варьирование повторностей; θ_5 — остаточное варьирование за счет факторов, неучтенных при наблюдениях.

1. Общая сумма квадратов отклонений находится по формуле

$$\theta = \sum_i x^2 - \frac{(\sum x)^2}{N}, \quad (4.37)$$

где θ — общая сумма квадратов отклонений; $\sum x^2$ — сумма квадратов всех вариантов, приведенных в табл. 4.18. В этой же таблице обозначены: S_1 — суммы вариант по столбцам; S_2 — суммы вариант по строкам; M_1 — средние арифметические по столбцам (годам); M_2 — средние арифметические по строкам (сортам).

$$\begin{aligned} \sum x^2 = 85 = \sum_1 x^2 = \sum_2 x^2 = 39 + 46 = 34 + 51 = 2^2 + 3^2 + \\ + 2^2 + 3^2 + 3^2 + 2^2 = 2^2 + 2^2 + 3^2 + 3^2 + 4^2 + 2^2. \end{aligned}$$

Возведем в квадрат сумму всех вариант:

$$(\sum x)^2 = 31^2 = 961.$$

Для контроля следует повторить вычисление общей суммы вариант по равенству: $\sum S_1 = \sum S_2 = \sum x = 15 + 16 = 14 + 17 = 31$, а также проверить равенство: $\sum x^2 = \sum_1 x^2 = \sum_2 x^2$, N — объем выборки равен 12.

По формуле (4.37) $\theta = 85 - (31^2/12) = 4,917$.

2. Вычисление остальных сумм квадратов, участвующих в равенстве (4.36), производится по одной в принципе общей формуле

$$\theta_i = \left[\sum S^2 - \frac{(\sum x)^2}{n} \right] / k, \quad (4.38)$$

где θ_i — сумма квадратов отклонений для какого-либо фактора; $\sum S^2$ — сумма квадратов сумм по градациям того фактора, для которого вычисляется θ_i ; n — число этих сумм, оно равно числу градаций, рассматриваемого фактора; k — число слагаемых (вариант), составляющих каждую сумму; $(\sum x)^2$ — квадрат суммы всех вариант.

В любом случае: $kn = N$, что можно использовать для проверки действий по правильному выбору сумм и числа их составляющих вариант.

Применяя конкретно формулу (4.38), вычислим сумму квадратов отклонений для фактора 1 (годы выращивания):

$$\theta_1 = \left[\sum S_1^2 - \frac{(\sum x)^2}{n_1} \right] / k_1,$$

где θ_1 — сумма квадратов отклонений по фактору 1; $\sum S_1^2$ — сумма квадратов сумм по градациям фактора 1, она равна $15^2 + 16^2 = 225 + 256 = 481$; $(\sum x)^2$ — квадрат суммы всех вариант, он равен $31^2 = 961$; n_1 — число сумм (градаций) фактора 1, оно равно 2 (число лет выращивания); k_1 — число слагаемых (вариант) в каждой отдельной сумме, оно равно 6.

Проверка: $n_1 \cdot k_1 = 2 \cdot 6 = 12$.

Сумма квадратов отклонений по первому фактору равна:

$$\theta_1 = \left(481 - \frac{31^2}{2} \right) / 6 = 0,083.$$

3. Сумма квадратов отклонений по второму фактору также вычисляется по формуле (4.38) следующим образом:

$$\theta_2 = \left[\sum S_2^2 - \frac{(\sum x)^2}{n_2} \right] / k_2,$$

где θ_2 — сумма квадратов отклонений фактора 2 (сорта); $\sum S_2^2$ — сумма квадратов сумм по грациям фактора 2, она равна: $14^2 + 17^2 = 196 + 289 = 485$; $(\sum x)^2 = 31^2 = 961$; n_2 — число сумм по фактору 2, оно равно здесь числу сортов, т. е. 2; k_2 — число вариантов, слагающих каждую из сумм S_2 , оно равно 6.

Проверка: $n_2 k_2 = N = 2 \cdot 6 = 12$.

Сумма квадратов отклонений по фактору 2:

$$\theta_2 = \left(485 - \frac{31^2}{2} \right) / 6 = 0,750.$$

4. Сумму квадратов отклонений для определения доли варьирования, вызываемой эффектом взаимодействия факторов 1 и 2, вычисляем по формуле:

$$\theta_3 = \left[\sum S_3^2 - \frac{(\sum x)^2}{n_3} \right] / k_3 - \theta_1 - \theta_2, \quad (4.39)$$

где θ_3 — сумма квадратов отклонений по эффекту совместного действия факторов 1 и 2; $\sum S_3^2$ — сумма квадратов сумм по отдельным клеткам комбинационной таблицы, в нашем случае она равна: $7^2 + 7^2 + 8^2 + 9^2 = 243$; $(\sum x)^2 = 31^2 = 961$; n_3 — число сумм по отдельным клеткам таблицы, $n_3 = 4$ (число «полей» четырехпольной таблицы); k_3 — число слагаемых в каждой из сумм S_3 , оно равно 3 (число повторностей); θ_1 , θ_2 — суммы квадратов факторов 1 и 2.

По формуле (4.39):

$$\theta_3 = \left(243 - \frac{31^2}{4} \right) / 3 - 0,083 - 0,750 = 0,084.$$

5. Сумму квадратов отклонений по повторностям определим по формуле (4.38) в следующем ее конкретном виде:

$$\theta_4 = \left[\sum S_4^2 - \frac{(\sum x)^2}{n_4} \right] / k_4,$$

где θ_4 — сумма квадратов отклонений по повторностям; $\sum S_4^2$ — сумма квадратов сумм по повторностям, их всего 3 суммы.

Суммы S_4 по 3 повторностям подсчитываются следующим образом. При каждом сорте в табл. 4.18 имеется 3 строки данных, соответствующих повторностям 1, 2 и 3 сверху вниз. Сумма S_4 по первой повторности будет равна: (2 + 2) по I сорту и (3 + 3) по II сорту,

т. е. $S_4 = (2 + 2) + (3 + 3) = 10$; по второй повторности: $S_4 = (3 + 2) + (3 + 4) = 12$ и по третьей повторности: $S_4 = (2 + 3) + (2 + 2) = 9$. Таким образом: $\Sigma S_4^2 = 10^2 + 12^2 + 9^2 = 325$, $(\Sigma x)^2 = 31^2 = 961$; n_4 — число сумм по повторностям, оно равно 3; k_4 — число слагаемых в каждой сумме, оно равно 4.

Проверка: $n_4 k_4 = N = 3 \cdot 4 = 12$.

В результате получим:

$$\theta_4 = \left[325 - \frac{31^2}{3} \right] / 4 = 1,167.$$

6. Сумму квадратов сумм отклонений по остаточному варьированию обычно определяют из равенства (4.36) следующим образом:

$$\theta_5 = \theta - \theta_1 - \theta_2 - \theta_3 - \theta_4,$$

т. е.: $\theta_5 = 4,917 - 0,083 - 0,750 - 0,084 - 1,167 = 2,833$.

Следует заметить, что в величину остаточной вариабельности θ_5 входит, кроме разнообразных случайных отклонений, также и вариабельность за счет взаимодействий факторов второго и выше порядков, которая обычно не вычисляется ввиду ее незначительности.

7. Для каждой из вычисленных сумм квадратов отклонений необходимо определить число степеней свободы.

Для θ оно равно общему числу вариантов без единицы:

$$v = N - 1 = 12 - 1 = 11.$$

Для θ_1 число степеней свободы равно числу градаций фактора 1 без единицы:

$$v_1 = n_1 - 1 = 2 - 1 = 1.$$

Для θ_2 число степеней свободы равно числу градаций фактора 2 без единицы:

$$v_2 = n_2 - 1 = 2 - 1 = 1.$$

Для θ_3 число степеней свободы равно произведению чисел степеней свободы по обоим факторам:

$$v_3 = v_1 v_2 = 1 \cdot 1 = 1.$$

Для θ_4 число степеней свободы равно числу повторностей без единицы:

$$v_4 = n_4 - 1 = 3 - 1 = 2.$$

Для θ_5 число степеней свободы равно:

$$v_5 = v - v_1 - v_2 - v_3 - v_4 = 11 - 1 - 1 - 1 - 2 = 6.$$

8. Дальнейшие вычисления удобнее вести по табл. 4.19, для 1, 2 и 3 столбцов которой данные уже получены. Дисперсии (столбец 4) вычисляются делением величин столбца 2 на соответствующие числа степеней свободы из столбца 3, например, $4,917/11 = 0,447$, $0,083/1 = 0,083$ и т. д.

Таблица 4.19. Результаты дисперсионного анализа по данным таблицы 4.18

Варьирование данных	Сумма квадратов отклонений	Степень свободы	Дисперсия, σ^2	Критерий Фишера	
				F	F' ($P'_1 = 0,95$)
1	2	3	4	5	6
Общее	$\theta = 4,917$	$\nu = 11$	0,447	1,06	3,09
По фактору 1	$\theta_1 = 0,083$	$\nu_1 = 1$	0,083	5,69	234,0
По фактору 2	$\theta_2 = 0,750$	$\nu_2 = 1$	0,750	1,59	5,99
За счет взаимодействия факторов 1 и 2	$\theta_3 = 0,084$	$\nu_3 = 1$	0,084	5,62	234,0
По повторностям	$\theta_4 = 1,167$	$\nu_4 = 2$	0,584	1,24	5,14
Остаточное	$\theta_5 = 2,833$	$\nu_5 = 6$	0,472	1,00	—

9. В столбце 5 вычисляем отношения каждой из дисперсий к остаточной, или наоборот. Если остаточная дисперсия меньше сравниваемой с ней, то она берется в знаменателе, т. е. в качестве делителя, а если остаточная дисперсия больше сравниваемой с ней, то остаточная дисперсия, наоборот, берется в качестве делимого, а сравниваемая с ней в качестве делителя.

10. Оценку полученных дисперсий производим по табл. 9П. В строке ν (1) отыскиваем число степеней свободы, соответствующее большей из сравниваемых дисперсий, а в столбце ν (2) находим число степеней свободы меньшей из двух сравниваемых дисперсий. На пересечении ν (1) и ν (2) в табл. 9П находится искомое значение F' . Например, по фактору 1 отношение дисперсий равно $0,472/0,083 = 5,69$. Для большей дисперсии число степеней свободы равно $\nu_5 = 6 = \nu$ (1), для меньшей: $\nu_1 = 1 = \nu$ (2). На пересечении ν (1) = 6 и ν (2) = 1 находим $F' = 234,0$. По фактору 2 при ν (1) = 1, ν (2) = 6, $F' = 5,99$. Если фактическое значение критерия Фишера F больше табличного F' , то действие данного фактора считается существенным. Если фактическое значение критерия меньше табличного, то действие изучаемого фактора на опытный объект считается недоказанным. В табл. 4.19 все фактические числа критерия (столбец 5) меньше табличных (столбец 6), поэтому делаем вывод о недостоверности влияния факторов 1 и 2, а также их взаимодействия на урожайность томата при заданных конкретных условиях. Дальнейший анализ данного опыта не имеет смысла, однако рассмотрим методику вычислений всех остальных показателей, которые могут быть полезны в других случаях.

11. Ошибка средней арифметической всего опыта в результате дисперсионного анализа определяется по формуле:

$$m_M = \sqrt{\frac{\sigma_{\text{ост}}^2}{N}}, \quad (4.40)$$

где m_M — ошибка опыта; $\sigma_{\text{ост}}^2$ — остаточная дисперсия, которая здесь равна 0,472 (табл. 4.19, столбец 4); N — объем выборки.

Для данных табл. 4.18 ошибка средней арифметической всего опыта:

$$m_M = \sqrt{\frac{0,472}{12}} = 0,198.$$

12. Ошибка частных средних арифметических по любым вариантам опыта также вычисляется по формуле (4.40), но в качестве численности группы берется та, по которой определялась средняя арифметическая:

$$m_M = \sqrt{\frac{\sigma_{\text{ост}}^2}{n}},$$

где m_M — ошибка частной средней арифметической; n — численность группы, по которой вычислена средняя арифметическая; $\sigma_{\text{ост}}^2$ — остаточная дисперсия.

Например, определим ошибку средних арифметических по сортам I и II: $m_M = \sqrt{0,472/6} = 0,280$.

Ошибка средних арифметических урожая по годам также равна 0,280, но это совпадение случайное, ввиду того что совпали численности групп данных, по которым определяются те и другие средние арифметические: по 6 чисел в каждой группе.

13. Ошибка разности любых средних арифметических из таблицы дисперсионного анализа находится по формуле:

$$m_\delta = \sqrt{\frac{2\sigma_{\text{ост}}^2}{n}}, \quad \text{или} \quad m_\delta = m_M \sqrt{2}, \quad (4.41)$$

где m_δ — ошибка разности средних арифметических; $\sigma_{\text{ост}}^2$ — остаточная дисперсия; n — численность меньшей из сравниваемых частных групп (здесь численность сравниваемых групп одинакова).

Для примера, рассматриваемого в этом параграфе, ошибка разности средних арифметических урожая и по сортам и по годам одинакова (численность групп совпадает):

$$m_\delta = \sqrt{\frac{2 \cdot 0,472}{6}} = 0,396.$$

Для оценки достоверности различия между частными средними арифметическими существуют различные критерии; наиболее простой из них критерий Стьюдента:

$$t = \frac{M_i - M_j}{m_\delta}, \quad (4.42)$$

где t — критерий Стьюдента, сравниваемый с табличным его значением по табл. ЗП при числе степеней свободы для остаточной дисперсии; M_i , M_j — сравниваемые средние арифметические; m_δ — ошибка разности средних арифметических, вычисляемая по формуле (4.41).

Сравним средние арифметические по урожайности сортов I и II (табл. 4.18):

$$t = \frac{2,83 - 2,33}{0,396} = 1,263.$$

По табл. 3П при $v = 6$ и $P_1 = 95\%$ $t = 2,447$. Поскольку вычисленное значение критерия (1,263) меньше табличного, считается, что различие по урожайности между сортами I и II недоказано.

Сравним урожаи по годам (табл. 4.18):

$$t = \frac{2,67 - 2,50}{0,396} = 0,429.$$

Сравнение этой величины с табличной также показывает отсутствие достоверного различия урожая по годам.

14. Для оценки качества опыта в целом можно руководствоваться во многих случаях показателем точности опыта, который вычисляется по формуле (§ 1.14):

$$P = \frac{m_M}{M} \cdot 100,$$

где P — показатель точности опыта; m_M — ошибка средней арифметической всего опыта, вычисляется по формуле (4.40), в нашем примере равна 0,198; M — средняя арифметическая всего опыта, вычисляется по формуле (1.08): $M = 31/12 = 2,58$.

Показатель точности опыта отсюда:

$$P = \frac{0,198}{2,58} \cdot 100 = 7,67\%.$$

Как известно из § 1.14, точность опыта считается неудовлетворительной, если $P > 5\%$.

15. Коэффициент вариации опытных данных вычисляется по формуле

$$v = \frac{\sqrt{\sigma^2}}{M} \cdot 100, \quad (4.43)$$

где v — коэффициент вариации; σ^2 — общая дисперсия (первое число в столбце 4 табл. 4.18); она равна: $\frac{\theta}{v} = \frac{4,917}{11} = 0,447$; M — средняя арифметическая опытных данных.

По формуле (4.43):

$$v = \frac{\sqrt{0,447} \cdot 100}{2,58} = 26\%.$$

Для получения более достоверных результатов в рассматриваемом примере следует заложить опыт в большем числе повторностей. Следует иметь в виду, что в двухфакторных опытах с многолетними растениями, растущими на одном месте, когда погодные условия года являются фактором, может возникнуть корреляция между градициями факторов опыта и почвенными условиями, поэтому последние необходимо выделять в качестве третьего фактора.

§ 4.17. Однофакторный дисперсионный анализ небольшой группы данных

Наиболее распространенным случаем при математической обработке данных полевого опыта является однофакторный дисперсионный анализ, когда изучается действие на объект опыта одного фактора (в различном числе его градаций).

Однофакторный дисперсионный анализ невозможно выполнить для опыта, проведенного без повторностей. Минимум повторностей — две, но это их число редко дает достоверные результаты; в полевых опытах обычно делают не меньше 3—4 повторностей. Рассмотрим, порядок проведения однофакторного дисперсионного анализа.

В табл. 4.20 приведены данные полевого опыта, в котором семена овса высевали на делянки с добавкой различных микроэлементов, в трех повторностях. Фактором, изучаемым здесь при помощи однофакторного дисперсионного анализа, является различие по содержанию микроэлементов в почве. Объектом наблюдения было среднее число колосков, заложившихся на главном генеративном побеге в результате влияния этого фактора. Число градаций действующего фактора — 4 (контроль и 3 градации опыта). Вычисления проводятся в следующем порядке.

1. В столбцах 2—4 табл. 4.20 приводятся опытные данные и тут же в скобках их квадраты. Например, 19,6 и $19,6^2 = (384,16)$.

2. Суммируем данные по строкам и запишем суммы S_1 в столбце 5: $19,6 + 12,6 + 10,8 = 43,0$ и т. д. Суммируем данные по столбцам и получаем суммы S_2 : $19,6 + 19,7 + 25,2 + 28,0 = 92,5$ и т. д. Суммы сумм по строкам и столбцам должны совпадать и давать сумму всех дат, т. е. $\sum x = 253,5$.

3. Аналогично этому суммируем квадраты всех дат (числа в скобках) по строкам ($\sum_1 x^2$) и по столбцам ($\sum_2 x^2$), суммы сумм которых также должны совпадать и давать $\sum x^2 = 6221,69$, т. е. сумму квадратов всех вариантов.

4. Возводим в квадрат все отдельные суммы S_1 и S_2 и затем получаем суммы их квадратов: $\sum S_1^2 = 18210,61$ и $\sum S_2^2 = 21555,47$.

5. Вычисляем средние арифметические по градациям фактора (столбец 8) и по повторностям (внизу столбцов 2, 3, 4). Средние по повторностям получаем делением каждой из сумм S_2 на 4 (число слагаемых этой суммы), а средние по вариантам опыта — делением сумм S_1 на 3 (число слагаемых каждой из сумм S_1). Например, средняя первой повторности равна $92,5/4 = 23,13$ (столбец 2), средняя по градациям фактора «контроль» равна $43,0/3 = 14,33$. Общая средняя арифметическая всех дат опыта равна:

$$(\sum x)/N = M = 253,5/12 = 21,13.]$$

6. Задачей однофакторного дисперсионного анализа является расчленение общего, или суммарного, варьирования объекта опыта на составные части по равенству:

$$\theta = \theta_1 + \theta_2 + \theta_3, \quad (4.44)$$

Таблица 4.20. Вычисление сумм квадратов для однофакторного дисперсионного анализа

Условия выра- вания (градации фактора)	Повторность			S_1	$\sum_1 x^2$	S_1^2	M_1
	I	II	III				
1	2	3	4	5	6	7	8
Контроль	19,6 (384,16)	12,6 (158,76)	10,8 (116,64)	43,0	659,56	1849,0	14,33
1 доза	19,7 (388,09)	9,3 (86,49)	10,2 (104,04)	39,2	578,62	1536,64	13,07
2 дозы	25,2 (635,04)	23,5 (552,25)	28,2 (795,24)	76,9	1982,53	5913,61	25,63
3 дозы	28,0 (784,00)	30,7 (942,49)	35,7 (1274,49)	94,4	3000,98	8911,36	31,47
S_2	92,5	76,1	84,9	$\sum x = 253,5$		$18210,61 = \sum S_1^2$	
$\sum_2 x^2$	2191,29	1739,99	2290,41	$\sum x^2 = 6221,69$			
S_2^2	8556,25	5791,21	7208,01	$\sum S_2^2 = 21555,47$			
M_2	23,13	19,03	21,23				

где θ — сумма квадратов отклонений по общему варьированию данных; θ_1 — сумма квадратов отклонений по грациям фактора опыта; θ_2 — сумма квадратов отклонений по повторностям опыта; θ_3 — сумма квадратов отклонений по остаточному варьированию данных опыта.

Общую сумму квадратов отклонений опытных данных находим по формуле (4.37):

$$\theta = \sum x^2 - \frac{(\sum x)^2}{N} = 6221,69 - \frac{253,5^2}{12} = 866,50.$$

7. Сумму квадратов отклонений по грациям условий опыта (факториальную) определяем по формуле (4.38):

$$\theta_1 = \left[\sum S_1^2 - \frac{(\sum x)^2}{n_1} \right] / k_1 = \left(18210,61 - \frac{253,5^2}{4} \right) / 3 = 715,02,$$

где $\sum S_1^2$ — сумма квадратов сумм по грациям опыта (см. итог столбца 7, табл. 4.20); $\sum x$ — сумма всех вариантов опыта (столбец 5); n_1 — число сумм S_1 , оно равно числу граций фактора опыта, т. е. 4; k_1 — число слагаемых в каждой из сумм S_1 , оно равно 3. Должно выдерживаться равенство: $n_1 k_1 = N$, т. е. $4 \cdot 3 = 12$.

8. Сумму квадратов отклонений по повторностям опыта вычисляем также по формуле (4.38):

$$\theta_2 = \left[\sum S_2^2 - \frac{(\sum x)^2}{n_2} \right] / k_2 = \left(21555,47 - \frac{253,5^2}{3} \right) / 4 = 33,68,$$

где θ_2 — сумма квадратов отклонений по повторностям опыта; $\sum S_2^2$ — сумма квадратов сумм по повторностям (см. внизу столбца 5); $\sum x$ — сумма всех вариантов; n_2 — число сумм, соответствующее числу повторностей, оно равно 3; k_2 — число слагаемых в каждой сумме S_2 , оно равно 4.

9. Сумма квадратов отклонений по остаточной вариации определяется из равенства (4.44): $\theta_3 = \theta - \theta_1 - \theta_2 = 866,50 - 715,02 - 33,68 = 117,80$. Таким образом, все составляющие уравнения (4.44) теперь известны.

10. Дальнейшие вычисления удобнее вести в табличной форме (табл. 4.21). Числа степеней свободы, приведенные в столбце 3, определяются следующим образом.

Число степеней свободы по общей сумме квадратов отклонений равно числу всех вариантов без единицы: $v = N - 1 = 12 - 1 = 11$, по грациям опыта (факториальной сумме квадратов) равно числу граций опыта без единицы: $v_1 = n_1 - 1 = 4 - 1 = 3$. Число степеней свободы для суммы квадратов отклонений по повторностям равно числу повторностей без единицы: $v_2 = n_2 - 1 = 3 - 1 = 2$; число степеней свободы остаточной суммы квадратов отклонений равно: $v_3 = v - v_1 - v_2 = 11 - 3 - 2 = 6$.

11. Дисперсии в столбце 4 — это частные от деления сумм квадратов отклонений на соответствующие им числа степеней свободы,

Таблица 4.21. Дисперсионный анализ данных табл. 4.20

Варьирование данных	Сумма квадратов отклонений	Степень свободы	Дисперсия, $\sigma_i^2 = \frac{\theta_i}{v_i}$	Критерий Фишера	
				F	F' ($P'_1=0,95$)
1	2	3	4	5	6
Общее	866,50 = θ	11 = v	78,77	—	—
По градациям опыта (факториальное)	715,02 = θ_1	3 = v_1	238,34	12,14	4,76
По повторностям	33,68 = θ_2	2 = v_2	16,84	1,17	19,33
Остаточное	117,80 = θ_3	6 = v_3	19,63	—	—

например:

$$\sigma^2 = 866,50/11 = 78,77; \quad \sigma_1^2 = 715,02/3 = 238,34 \text{ и т. д.}$$

12. Полученные дисперсии сравниваются с остаточной дисперсией, которая равна: $\sigma_{\text{ост}}^2 = 19,63$. При этом бóльшая дисперсия всегда делится на меньшую (столбец 5):

$$F_1 = \frac{\sigma_1^2}{\sigma_{\text{ост}}^2} = \frac{238,34}{19,63} = 12,14; \quad F_2 = \frac{\sigma_{\text{ост}}^2}{\sigma_2^2} = \frac{19,63}{16,84} = 1,17.$$

Вычисленные значения критерия Фишера сравниваем с табличными (по табл. 9П), которые на $P'_1 = 0,95$ уровне достоверности приведены в столбце 6 табл. 4.21 (о пользовании табл. 9П см. § 4.16, пункт 10). Сравнение вычисленных величин критерия с табличными показывает, что содержание микроэлементов в почве достоверно влияет на число заложившихся колосков у овса, так как $F_1 > F'_1$, т. е. $12,14 > 4,76$.

13. Ошибка общей средней арифметической определяется по формуле (4.40):

$$m_M = \sqrt{\frac{19,63}{12}} = 1,23.$$

Таким образом, общая средняя арифметическая равна: $M = 21,13 \pm 1,28$.

14. Ошибка частных средних арифметических определяется по формуле (4.40) с той разницей, что здесь вместо N — общей численности вариант, подставляется n — численность группы, для которой вычислена частная средняя арифметическая; в частности ошибка средних по повторностям равна:

$$m_M = \sqrt{\frac{\sigma_{\text{ост}}^2}{n}} = \sqrt{\frac{19,63}{4}} = 2,21,$$

т. е. средние арифметические (см. табл. 4.20) равны: $23,13 \pm 2,21$; $19,03 \pm 2,21$; $21,23 \pm 2,21$.

Ошибка средних арифметических по грациям фактора (по условиям выращивания) равна:

$$m_M = \sqrt{19,63/3} = 2,56.$$

Отсюда средние арифметические по грациям условий выращивания (см. табл. 4.20) равны: $14,33 \pm 2,56$; $13,07 \pm 2,56$; $25,63 \pm 2,56$; $31,47 \pm 2,56$.

15. Ошибки разностей средних арифметических, при помощи которых оценивается достоверность различия последних, вычисляются по формуле (4.41). Ошибка разностей средних арифметических по грациям фактора: $m_\delta = \sqrt{2} \cdot 2,56 = 3,62$. Ошибка разности средних арифметических по повторностям: $m_\delta = \sqrt{2} \cdot 2,21 = 3,13$.

16. Величину различий средних арифметических при помощи ошибок разностей оценим по критерию Стьюдента (4.42). Сравним средние арифметические по грациям фактора, т. е. по условиям выращивания, с контролем:

$$t_1 = (14,33 - 13,07)/3,62 = 0,35;$$

$$t_2 = (25,63 - 14,33)/3,62 = 3,12;$$

$$t_3 = (31,47 - 14,33)/3,62 = 4,73.$$

Полученные значения критерия сравниваются с табличным, которое при числе степеней свободы $\nu = 6$ (для остаточной дисперсии) на доверительном уровне $P'_1 = 0,95$ равно $t' = 2,447$ (табл. 3П). Существенно отличаются от контроля средние по 2 и 3 грациям опыта, условия которых, следовательно, достоверно влияют на заложение колосков у овса.

Таким же образом, по формуле (4.42) можно оценить различие средних и по повторностям при помощи ошибки их разностей, равной $m_\delta = 3,13$, и также при $\nu = 6$. Сравнение средних по повторностям полезно для проверки однородности последних.

17. Ошибки общей средней арифметической по грациям фактора и по повторностям можно использовать для вычисления трех соответствующих показателей точности опыта, по формуле (1.50) из § 1.14. Показатель точности опыта для общей средней арифметической равен:

$$P = (1,28/21,13) \cdot 100 = 6,06 \% \quad (\text{см. данные в пунктах 5 и 13}).$$

Показатель точности опыта для частных средних арифметических по грациям фактора вычислим только для наименьшей средней арифметической (см. выше пункт 14):

$$P = (2,56/13,07) \cdot 100 = 19,59 \%.$$

Показатель точности опыта для частных средних по повторностям также вычислим для наименьшей средней (см. выше пункт 14):

$$P = (2,21/19,08) \cdot 100 = 11,61 \%.$$

Наибольшее значение для оценки качества опыта в данном случае имеет показатель точности опыта, вычисленный для граций

фактора. Поскольку в рассматриваемом примере оценка критерием Фишера показала достоверность влияния условий выращивания на заложение колосков, величину показателя точности опыта $P = 19,59 > 5\%$ можно рассматривать как указание на желательное увеличение повторностей опыта и улучшение выровненности агрофона делянок.

§ 4.18. Двухфакторный дисперсионный анализ небольшой группы данных без повторностей

Нередки случаи, когда в ходе эксперимента фиксировались только однократные наблюдения по каждой совместной градации двух сопряженных факторов. Вследствие этого опыт, в котором изучалось совместное действие двух факторов на объект, не будет иметь повторностей, что иногда служит препятствием для математической обработки данных, без которой последние теряют свою научную ценность.

В подобных случаях может оказаться полезной предлагаемая ниже схема двухфакторного дисперсионного анализа без повторностей. Она принципиально не отличается от такой схемы однофакторного дисперсионного анализа, в которой из общей суммы квадратов особо выделяется варьирование, вызванное различием данных по повторностям (см. § 4.17). Разница между этими двумя схемами заключается в том, что в схеме, приводимой в § 4.18, градации повторности принимаются за градации второго фактора.

Рассматриваемый в этом параграфе метод рекомендуется применять тогда, когда данные уже получены и повторять эксперимент нецелесообразно или невозможно. В прочих случаях лучше планировать двухфакторный эксперимент таким образом, чтобы данные имели не менее 3—4 повторностей и их можно было бы обработать методами из § 4.16 или § 4.19 двухфакторного дисперсионного анализа с повторностями, что существенно повысит надежность получаемых выводов. От двухфакторного анализа с повторностями метод из § 4.18 отличается тем, что посредством него невозможно выявить ту долю варьирования данных, которая возникает из-за взаимодействия обоих факторов, ибо она объединяется с остаточной дисперсией, что обязательно следует учитывать при толковании результатов анализа.

Для примера выполним дисперсионный анализ без повторностей, выявляющий достоверность влияния предшествующей культуры на урожай последующей, по данным Госсортосети (табл. 4.22). Одним из факторов, влияющих на урожай пшеницы, является сортовая специфика последней, а другим фактором — предшествующая культура. Ввиду того что каждая совместная градация опыта содержит только одну оценку урожайности, данный дисперсионный анализ является двухфакторным без повторностей; проводится он в следующем порядке.

1. Как уже говорилось выше, все вычисления по двухфакторному анализу без повторностей можно проводить по схеме из § 4.17.

Таблица 4.22. Влияние предшествующей культуры и сортовой специфики пшеницы на ее урожай

Фактор 1 — сорт пшеницы	Фактор 2 — предшественник				S_1	S_1^2	M_1
	Кукуруза	Яровая пшеница	Бобы комовые	Сахарная свекла			
1	2	3	4	5	6	7	8
1. Лютесценс 758	19,4	19,8	11,1	10,8	61,1	3733,21	15,27
2. Мильтурум 553	17,6	19,7	9,4	9,1	55,8	3113,64	13,95
3. Харьковская 46	10,9	11,8	9,9	10,2	42,8	1831,84	10,70
4. S_2	47,9	51,3	30,4	30,1	159,7	8678,69	
5. S_2^2	2294,41	2631,69	924,16	906,01	6756,27		
6. M_2	15,97	17,10	10,13	10,03			$M = 13,31$

Целью дисперсионного анализа без повторностей также является разложение общей дисперсии на составляющие ее части:

$$\theta = \theta_1 + \theta_2 + \theta_3,$$

где θ — сумма квадратов отклонений по общему варьированию; θ_1 — сумма квадратов отклонений по грациям первого фактора; θ_2 — сумма квадратов отклонений по грациям второго фактора; θ_3 — сумма квадратов отклонений остаточной дисперсии.

2. Составим комбинационную табл. 4.22 и выполним указанные в ней действия. Сумма всех вариантов $\Sigma x = 159,7$. В столбце 6 приведены суммы вариантов по строкам (S_1), а в строке 4 — суммы вариантов по столбцам (S_2). В столбце 7 (S_1^2) и строке 5 (S_2^2) эти суммы возведены в квадрат и квадраты суммированы. В столбце 8 и в строке 6 вычислены средние арифметические по грациям факторов, например по предшественнику — кукурузе:

$$M_2 = (19,4 + 17,6 + 10,9)/3 = 15,97.$$

Средний урожай сорта Лютесценс 758 равен:

$$M_1 = (19,4 + 19,8 + 11,1 + 10,8)/4 = 15,27.$$

Общая средняя арифметическая урожая в опыте:

$$M = (\Sigma x)/N = 159,7/12 = 13,31.$$

3. При многозначных вариантах сумму квадратов отклонений по общему варьированию данных целесообразно вычислять через разности всех вариантов с условным числом A (в данном случае прием $A = 10$), затем возвести разности в квадрат и суммировать их.

В табл. 4.23 в первых трех строках (1—3) первых четырех столбцов (1—4) приведены разности $x - 10$. В строке 4 и столбце 5 при-

Таблица 4.23. Вычисление суммы квадратов данных из табл. 4.22 (объяснения см. в тексте) $A = 10$

Номер строки	1	2	3		5
1	9,4	9,8	1,1	0,8	21,1
2	7,6	9,7	-0,6	-0,9	15,8
3	0,9	1,8	-0,1	0,2	2,8
4	17,9	21,3	0,4	0,1	39,7
5	88,36	96,04	1,21	0,64	186,25
6	57,76	94,09	0,36	0,81	153,02
7	0,81	3,24	0,01	0,04	4,10
8	146,93	193,37	1,58	1,49	343,37

ведены суммы разностей по строкам и столбцам (для контроля вычислений). В строках 5—7 столбцов 1—4 приведены квадраты соответствующих разностей. В строке 8 и столбце 5 — суммы квадратов разностей по строкам и столбцам (для контроля).

$$\Sigma (x - A) = 39,7; \Sigma (x - A)^2 = 343,37; [\Sigma (x - A)]^2 = 39,7^2 = 1576,09.$$

Сумма квадратов отклонений по общей вариабельности данных равна:

$$\theta = \Sigma (x - A)^2 - \frac{[\Sigma (x - A)]^2}{N} = 343,37 - \frac{39,7^2}{12} = 212,03.$$

4. Остальные вычисления можно делать, руководствуясь схемой из § 4.23, в которой для вычисления сумм квадратов отклонений применяется общая формула (4.38). Сумму квадратов отклонений по фактору 1 находим по данным из табл. 4.22: $\Sigma S_1^2 = 8678,69$; $(\Sigma x)^2 = 159,7^2 = 25504,09$; $n_1 = 3$; $k_1 = 4$; $n_1 k_1 = N$; $3 \cdot 4 = 12$; отсюда (см. § 4.17, пункт 7):

$$\theta_1 = \left(8678,69 - \frac{159,7^2}{3} \right) / 4 = 44,33.$$

5. Сумма квадратов отклонений по фактору 2 по данным из табл. 4.20 определится из выражения (см. § 4.17, пункт 8):

$$\theta_2 = \left[\Sigma S_2^2 - \frac{(\Sigma x)^2}{n_2} \right] / k_2 = \left(6756,27 - \frac{159,7^2}{4} \right) / 3 = 126,75.$$

6. Сумма квадратов отклонений остаточной дисперсии определится из равенства:

$$\theta_3 = \theta - \theta_1 - \theta_2 = 212,03 - 44,33 - 126,75 = 40,95.$$

7. Дальнейшие вычисления удобнее вести в форме табл. 4.24, где приводятся следующие данные. Суммы квадратов отклонений по источникам варьирования из столбца 2 делят на соответствующие числа степеней свободы из столбца 3, результаты деления (диспер-

Таблица 4.24. Результаты двухфакторного дисперсионного анализа без повторностей по данным табл. 4.22

Вариабельность данных	Сумма квадратов отклонений	Степень свободы, ν	Дисперсия, σ^2	Критерий Фишера	
				F	$F' (P'_1 = 0,95)$
1	2	3	4	5	6
Общая	212,03	11	19,27	2,83	—
По градациям первого фактора	44,33	2	22,16	3,25	5,14
По градациям второго фактора	126,75	3	42,25	6,19	4,76
Остаточная	40,95	6	6,82	—	—

сии) записаны в столбце 4. Числа степеней свободы определяются по тому же правилу, что и в § 4.17, пункт 10.

Дисперсии в столбце 4 для получения критерия Фишера сравнивают с остаточной дисперсией, которая равна 6,82. При этом всегда большая дисперсия делится на меньшую. В данном случае все дисперсии больше остаточной, поэтому последняя является делителем:

$$19,27/6,82 = 2,83; \quad 22,16/6,82 = 3,25; \quad 42,25/6,82 = 6,19.$$

Полученные величины критерия (столбец 5) сравнивают с табличными величинами (табл. 9П). При $\nu(1) = 3$ и $\nu(2) = 6$ $F' = 4,76$. При $\nu(1) = 2$, $\nu(2) = 6$ $F' = 5,14$. Поскольку 3,25 меньше, чем 5,14, делаем вывод о том, что разница в сорте (фактор 1) у пшеницы в данных условиях не вызывает существенного влияния на урожай. Поскольку 6,19 больше, чем 4,76, делаем вывод о том, что различие в видах предшественников (фактор 2) существенно влияет на урожай. Как указывалось в начале § 4.18, по данному методу невозможно выделить из остаточной дисперсии ту долю варьирования, которая вызвана взаимодействием обоих факторов. Различие в предшественниках, по-видимому, больше влияет на урожай пшеницы, чем разница в сортах у последней.

§ 4.19. Двухфакторный дисперсионный анализ большой группы данных

В § 4.16 была изложена методика двухфакторного анализа с повторностями и в § 4.22 — дисперсионного анализа без повторностей, в обоих случаях для небольшого числа дат. Принципы обработки двухфакторным анализом больших выборок остаются теми же, однако техника вычислений при этом несколько изменяется. При обработке данных методом из § 4.19 важно правильно составить исходную комбинационную таблицу. Для этого желательно при планировании эксперимента предусматривать последующее применение

Таблица 4.25. Комбинационная таблица двухфакторного дисперсионного анализа данных по влиянию на высоту растений числа соцветий (фактор 1) и диаметра соцветий (фактор 2) у гелениума

Номер строки	Высота растений, см	1. Градации по числу соцветий										Накопленные частоты			
		72—158					159—245							246—333	
		2. Градации по диаметру соцветий													
		Классы		y	4,0—4,9		5,0—6,0		4,0—4,9		5,0—6,0		q ₁ = 70	q ₂ = 15	
		4	5		6	7	8	9	10	11	12				
	1	2	3	4	5	6	7	8	9	10	11	12			
1	111—128	120	+2												
2	93—110	102	+1	5	14	2	10	3		15	15	15			
3	75—92	84	0	9	14	6	8	4		40	55	—			
4	57—74	66	-1	4	7	4	17	10		59	—	—			
5	39—56	48	-2		10	2		3		16	26	—			
6	f _x			18	45	14	35	8	20	140	q ₃ = 36	q ₄ = 10			
7	∑ f _{xy} ^a			+1	-13	+8	+28	+3	+7	+34					
8	(∑ f _{xy} ^a) ²			1	169	64	784	9	49	—					
9	$\frac{(\sum f_{xy}^a)^2}{f_x}$			0,055	3,756	4,571	22,400	1,125	2,450	34,357					

дисперсионного анализа, который дает возможность получить достоверные выводы, оцененные объективными критериями.

В рассматриваемом ниже примере исследуется влияние на высоту стеблей у гелениума двух факторов: числа соцветий на одном растении — фактор 1 и диаметра соцветий — фактор 2 (табл. 4.25).

В двухфакторном дисперсионном анализе необходимо соблюдать условие о независимости между собой обоих факторов, действующих на рассматриваемый объект. С этой целью были вычислены показатели корреляции рангов по Спирмэну между высотой растений (x), числом соцветий на них (y) и диаметром соцветий (z). Каждый из этих рядов представлял собой 15 средних арифметических (по числу сортов гелениума).

Отсутствует корреляционная связь между высотой растений и диаметром соцветий $\rho_{xz} = 0,21$, а также между числом соцветий и их диаметром: $\rho_{yz} = 0,07$. Имеется корреляционная связь только между числом соцветий и высотой стеблей: $\rho_{xy} = 0,56$, однако последняя является не фактором, а объектом опыта.

Число данных по градациям должно быть равным, или же их отношение должно быть одинаковым. В табл. 4.25 суммы частот f_x в столбцах 4 и 5, 6 и 7, 8 и 9 находятся в одном отношении:

$$45/18 = 2,5; \quad 35/14 = 2,5; \quad 20/8 = 2,5.$$

По фактору 1 установлено 3 градации по числу соцветий: 72—158, 159—245 и 246—333. В пределах каждой из этих групп растений выделяется 2 одинаковые градации второго фактора по диаметру соцветий: 4,0—4,9 и 5,0—6,0 мм. Далее выполнены следующие действия.

1. По высоте растений примем 5 классов, убывающих по величине сверху вниз (столбец 1) и вычислим их середины y (столбец 2).

2. В столбце 3 записаны отклонения от условной средней $A_y = 84$ см по формуле,

$$a_y = \frac{y - A_y}{c},$$

где c — классовый интервал, равный здесь 18 см.

3. Согласно установленным градациям факторов 1, 2 и классам высоты растений разносим данные измерений по ячейкам таблицы (строки 1—5, столбцы 4—9). В итоговой строчке 6 и столбце 10 частоты суммированы. Обе суммы должны совпасть и дать общее число измерений:

$$N = 140 = \sum f_x = \sum f_y.$$

4. В строке 7 записываем суммы произведений частот f_{xy} из всех ячеек на отклонения a_y из столбца 3. Например: $5 \cdot (+1) + 9 \cdot 0 + 4 \cdot (-1) = +1$; $14 \cdot (+1) + 14 \cdot 0 + 7 \cdot (-1) + 10 \cdot (-2) = -13$ и т. д.

5. Квадраты чисел из строки 7 записываем в строке 8. Например: $(+1)^2 = 1$; $(-13)^2 = 169$ и т. д.

6. В строке 9 приведены частные от деления чисел из строки 8 на соответствующие им частоты из строки 6. Например: $1 : 18 = 0,055$; $169 : 45 = 3,756$ и т. д. Сумма этих частных равна 34,357.

7. В столбцах 11 и 12 приведены два ряда накопленных частот от исходных частот из столбца 10. Методика их получения и проверки изложена в § 1.10. Полусуммы $q_1 - q_4$ необходимы для получения сумм \sum_1 и \sum_2 , которые равны:

$$\sum_1 = q_1 - q_3 = 70 - 36 = 34,$$

$$\sum_2 = q_1 + q_3 + 2(q_2 + q_4) = 70 + 36 + 2(15 + 10) = 156.$$

Проверка вычисления \sum_2 производится подстановкой ее в равенство:

$$\sum f_y a^2 = \sum_2 = 15(+2)^2 + 40(+1)^2 + 59(0)^2 + 16(-1)^2 + 10(-2)^2 = 156.$$

Следовательно, вычисления сделаны верно.

Для проверки вычисления \sum_1 можно воспользоваться тем фактом, что сумма чисел строки 7 также должна быть равна \sum_1 , т. е. в данном случае +34.

8. Задача двухфакторного дисперсионного анализа, как и в предыдущих параграфах, заключается в разложении общей суммы квадратов отклонений на ее составляющие по равенствам:

$$\theta = \theta_x + \theta_z, \quad \theta_x = \theta_1 + \theta_2 + \theta_{1.2}, \quad (4.45)$$

где θ — общая сумма квадратов отклонений; θ_x — сумма квадратов отклонений факториальной дисперсии; θ_z — сумма квадратов отклонений остаточной дисперсии; θ_1 — сумма квадратов отклонений дисперсии первого фактора; θ_2 — сумма квадратов отклонений дисперсии второго фактора; $\theta_{1.2}$ — сумма квадратов отклонений дисперсии, вызванной взаимодействием первого и второго факторов.

Общую сумму квадратов отклонений (θ) находим по формуле

$$\theta = \sum_2 - \frac{\sum_1^2}{N}, \quad (4.46)$$

где \sum_1 , \sum_2 — суммы, вычисление которых приведено выше, в пункте 7; N — объем выборки.

Подставляя данные, получим по формуле (4.46):

$$\theta = 156,0000 - 34^2/140 = 147,743.$$

Заметим для дальнейших расчетов, что

$$\frac{\sum_1^2}{N} = \frac{34^2}{140} = 8,257.$$

9. Сумма квадратов отклонений факториальной дисперсии определится по формуле

$$\theta_x = \sum_x \frac{(\sum f_{xy} a_y)^2}{f_x} - \frac{\sum_1^2}{N}, \quad (4.47)$$

Таблица 4.26. Вычисление сумм квадратов отклонений дисперсий отдельных факторов (к таблице 4.25)

Номер ра строк	Градации факторов 1 и 2	f_x	$\Sigma f_{xy}a_y$	$(\Sigma f_{xy}a_y)^2$	$\frac{(\Sigma f_{xy}a_y)^2}{f_x}$	
	1	2	3	4	5	
1	1. Число соцветий	72—158	63	+1—13 = —12	(—12) ² = 144	2,286
2		159—245	49	+8+28 = +36	36 ² = 1296	26,449
3		246—333	28	+3+7 = +10	10 ² = 100	3,571
4		Сумма	140	+34	—	32,306
5	2. Диаметр соцветий	4,0—4,9	40	+1+8+3 = +12	(+12) ² = 144	3,600
6		5,0—6,0	100	—13+28+7 = +22	(+22) ² = 484	4,840
7		Сумма	140	+34	—	8,440

где первый член правой части уравнения представляет собой сумму чисел строки 9 табл. 4.25 и равен 34,357, а второй член вычислен в предыдущем пункте и равен 8,257. Итак:

$$\theta_x = 34,357 - 8,257 = 26,100.$$

10. Сумма квадратов отклонений остаточной дисперсии определится из равенства:

$$\theta_z = \sum_{j_2}^J - \sum_{j_1}^J \frac{(\sum f_{xy}a_y)^2}{f_x}, \quad (4.48)$$

где \sum_{j_2} — сумма, найденная по способу накопленных частот (см. пункт 7, выше), а второй член правой части уравнения представляет собой сумму чисел строки 9 из табл. 4.25. Отсюда: $\theta_z = 156,000 - 34,357 = 121,643$.

11. Сумму квадратов факториальной дисперсии θ_x необходимо расчленить на составляющие согласно равенству (4.45). Для этого определяем сумму квадратов отклонений по факторам 1 и 2 отдельно по формуле (4.47). В табл. 4.26 выполнены требуемые для этого действия, где числа в столбце 2 взяты из строки 6 табл. 4.25, т. е. $18 + 45 = 63$; $14 + 35 = 49$; $8 + 20 = 28$, а числа в столбце 3 — из строки 7 табл. 4.25.

По формуле (4.47):

$$\theta_1 = 32,306 - 8,257 = 24,049; \theta_2 = 8,440 - 8,257 = 0,183,$$

Таблица 4.27. Результаты двухфакторного дисперсионного анализа
(1 фактор — число соцветий, 2 фактор — диаметр соцветий на тех же
особях гелениума)

Варьирование данных	Сумма квадратов отклонений, θ_i	Степень свободы, ν_i	Дисперсия, $\sigma_i^2 = \frac{\theta_i}{\nu_i}$	Критерий Фишера из табл. 9П	
				F — фактическая величина	F' — табличное значение при $P'_1 = 0,95$
1	2	3	4	5	6
Общее	$\theta = 147,743$	$140 - 1 = 139$	1,063	—	—
Факториальное	$\theta_x = 26,100$	$3 \cdot 2 - 1 = 5$	5,222	5,75	2,21
В том числе					
по фактору 1	$\theta_1 = 24,049$	$3 - 1 = 2$	12,025	13,24	3,00
по фактору 2	$\theta_2 = 0,183$	$2 - 1 = 1$	0,183	4,96	254,3
За счет взаимодействия факторов 1 и 2	$\theta_{1,2} = 1,868$	$2 \cdot 1 = 2$	0,934	1,03	3,00
Остаточное	$\theta_z = 121,643$	$140 - 3 \cdot 2 = 134$	0,908	—	—

где 32,306 и 8,440 — итоги 4 и 7 строк табл. 4.26, а $8,257 = \Sigma_1^2/N$ (см. пункт 8).

Сумму квадратов отклонений дисперсии, вызванной совместным действием факторов 1 и 2 ($\theta_{1,2}$), найдем из равенства (4.45):

$$\theta_{1,2} = \theta_x - \theta_1 - \theta_2 = 26,100 - 24,049 - 0,183 = 1,868.$$

12. Дальнейшие вычисления лучше производить в форме табл. 4.27. Степени свободы в столбце 3 этой таблицы определяются следующим образом.

Для общей дисперсии число степеней свободы равно объему выборки без единицы: $\nu = N - 1$.

Для факториальной дисперсии: $\nu_x = k_1 k_2 - 1$, т. е. произведение числа градаций первого фактора (k_1) на число градаций второго фактора (k_2) без единицы, в том числе по дисперсии первого фактора: $\nu_1 = k_1 - 1$, т. е. числу градаций первого фактора без единицы, и по дисперсии второго фактора: $\nu_2 = k_2 - 1$ — числу градаций второго фактора без единицы.

Для дисперсии за счет взаимодействия факторов: $\nu_{1,2} = \nu_1 \nu_2$.

Для остаточной дисперсии: $\nu_z = N - k_1 k_2$, т. е. объему выборки минус произведение числа градаций первого и второго факторов.

В столбце 4 табл. 4.27 вычислены дисперсии, т. е. частные от деления чисел столбца 2 на соответствующие им числа из столбца 3. Сравнение дисперсий с остаточной, которая равна $\sigma_z^2 = 0,908$, производится делением на нее всех сравниваемых дисперсий в столбце 4, за исключением дисперсии по $\theta_2 = 0,183$, которая оказалась

меньше, чем остаточная, и поэтому послужила делителем при сравнении. Результаты этого сравнения приведены в столбце 5. Табличные значения критерия Фишера в столбец 6 выписаны из табл. 9П, способ пользования которой указан в § 4.16, пункт 10.

Сравнивая фактические и табличные значения критерия Фишера, делаем вывод о том, что на высоту растений у гелениума осеннего оказывает достоверное влияние число соцветий ($13,83 > 3,00$) и отсутствует влияние диаметра соцветий ($4,96 < 254,3$).

Большое влияние на высоту растений оказывают другие неучтенные факторы, о чем свидетельствует значительная величина остаточной дисперсии. Очевидно, число соцветий, несмотря на достоверность влияния, не является основным фактором из тех, от которых зависит высота растений.

§ 4.20. Критерий Бартлетта

Для того чтобы оценить некоторый массив, состоящий из неодинаковых по численности групп данных по однородности их варьирования в целом, не обязательно прибегать к дисперсионному анализу; если не имеет значения, какие конкретно из групп отличаются по своей вариабельности от других, то с этой целью можно применить критерий Бартлетта. Основное условие его применения заключается в том, что данные должны иметь нормальное распределение.

Рассмотрим технику расчета критерия Бартлетта на примере из фенологии. При изучении причин различий в датах начала цветения за разные годы у 360 видов древесных растений в Ленинграде было предположено, что на сроки начала цветения воздействуют специфичные комплексы погоды каждого года. С этой целью решено при помощи критерия Бартлетта оценить варьирование дат цветения за 7 лет наблюдений: с 1956 по 1962 г. (табл. 4.28). В столбце 1 таблицы приведены даты начала цветения (в днях от 1 марта — x_0), которые для упрощения расчетов в столбце 2 таблицы переведены в натуральный ряд чисел по формуле:

$$x = \frac{x_0 - A}{c} = \frac{x_0 - 35}{33},$$

где A — первое число в ряду вариантов x_0 , c — классовый интервал. В следующих столбцах таблицы приведено число видов, зацветающих в те или иные сроки по годам; g — число лет наблюдений; n — сумма вариантов (видов) по столбцам, $1/(n - 1) \cdot 1000$ — числа, обратные сумме видов без одного и для сокращения записи умноженные на 1000, их сумма (0,0215936) приведена без умножения на 1000. Средняя арифметическая по каждому году (M) вычислена по формуле (1.13), а дисперсии через суммы квадратов отклонений от средней (см. § 1.10). Натуральные логарифмы величин дисперсий умножены далее на число наблюдаемых видов по каждому году без одного, после чего получены все исходные данные для вычисления

Таблица 4.28. К вычислению критерия Барлетта

Даты начала цветения от 1 марта x_0	$x = \frac{x_n - 35}{33}$	56	57	58	59	60	61	62	Σ
1	2	3	4	5	6	7	8	9	10
35	0	0	5	3	7	4	4	12	
68	1	53	118	30	119	86	85	95	
101	2	206	152	207	139	190	194	171	
134	3	37	51	61	40	38	31	65	
167	4	8	9	14	7	5	8	15	
200	5	0	2	4	1	0	0	2	
233	6	1	1	1	0	1	0	0	
$c = 33$	n	305	338	320	313	324	322	360	2282
	$1000 \cdot 1/(n - 1)$	3,2895	2,9674	3,1348	3,2051	3,096	3,1153	2,7855	0,0215936
	M	2,0131	1,855	2,2156	1,7572	1,8704	1,8572	1,95	
	$\Sigma (x - M)^2 f$	1236,1	1163,1	1570,7	966,49	1133,5	1110,7	1368,9	8549,49
	σ^2	4,0662	3,4514	4,9238	3,0977	3,5093	3,4602	3,8131	
	$\ln \sigma^2$	1,4027	1,2387	1,594	1,1307	1,2555	1,2413	1,3384	
	$(n - 1) \ln \sigma^2$	426,421	417,442	508,486	352,778	405,526	398,457	480,486	2989,596

критерия Бартлетта:

$$b = \left(\ln \frac{\sum \sum (x - M)^2 f}{\sum (n - 1)} \right) \cdot \sum (n - 1) - \sum \ln \sigma^2 (n - 1), \quad (4.49)$$

$$a = 1 + \frac{1}{3(g - 1)} \left(\sum \frac{1}{n - 1} - \frac{1}{\sum (n - 1)} \right),$$

$$\begin{aligned} b &= \left(\ln \frac{8549,49}{2275} \right) \cdot 2275 - 2989,596 = \\ &= 1,32389 \cdot 2275 - 2989,596 = 22,281; \end{aligned}$$

$$a = 1 + \frac{1}{3(7 - 1)} (0,0215936 - 0,00043956) = 1,001175.$$

Сумму $\sum (n - 1)$ можно найти из выражения: $\sum (n - 1) = \sum n - g = 2282 - 7 = 2275$, т. е. она равна сумме наблюдений по всем группам минус число групп.

Величина критерия Бартлетта равна:

$$b/a = 22,281/1,001175 = 22,255.$$

По табл. 10П при числе степеней свободы: $\nu = g - 1 = 7 - 1 = 6$ величина хи-квадрат на уровне значимости 0,05 равна 12,6, а на уровне значимости 0,001 она равна 22,5. Вычисленная величина хи-квадрат находится между этими двумя табличными величинами, т. е. на первом уровне значимости группы массива (годы) различаются между собой по вариабельности, а на втором уровне они не различаются. Следовательно, вопрос о различии лет по комплексу факторов погоды, влияющих на сроки начала цветения, остается открытым.

Приведенную схему расчета критерия Бартлетта можно применять и к невзвешенным рядам данных, а также при равночисленных сравниваемых группах, т. е. когда $n_1 = n_2 = \dots = n_g$. В последнем случае вычисления несколько упрощаются.

Рассмотрим еще один пример применения критерия Бартлетта к сравнению невзвешенных рядов одинакового объема. В табл. 4.29 в столбцах x приведены (в мм) величины диаметра соцветий у растений гелениума осеннего сорт Ротхаут, измеренные на однолетних растениях в течение трех лет.

Как показало биометрическое исследование этих данных, их распределение соответствует нормальному закону, что является условием применения критерия Бартлетта. Требуется выявить, различаются ли величины диаметров соцветий гелениума в зависимости от года выращивания. Цель расчетов в табл. 4.29, как и всегда при вычислении критерия Бартлетта, — получить величины сумм квадратов отклонений от средних и дисперсий по каждому из сравниваемых рядов. Каким способом они будут вычислены для применения формул (4.49) и (4.50), не имеет значения.

В данном случае суммы квадратов отклонений и дисперсии удобнее вычислить по способу из § 1.16. Сначала находим сумму величин по каждому ряду: 600; 576; 569, затем средние: $M = (\sum x)/n =$

Таблица 4.29. Критерий Бартлетта для равночисленных невзвешенных рядов

Первый год			Второй год			Третий год		
		a^2	x	a		x	a	a^2
62	2	4	54	-4	16	55	-2	4
58	-2	4	59	1	1	58	1	1
60	0	0	60	2	4	56	-1	1
60	0	0	44	-14	196	60	3	9
60	0	0	65	7	49	60	3	9
61	1	1	50	-8	64	59	2	4
60	0	0	63	5	25	59	2	4
61	1	1	60	2	4	54	-3	9
60	0	0	56	-2	4	50	-7	49
50	-2	4	65	7	49	58	1	1
n	10		10			10		
Σx	600		576			569		
M	60		57,6			56,9		
A	60		58			57		
Σa	4-4=0		24-28=-4			12-13=-1		
$(\Sigma a)^2$	0		16			1		
Σa^2	14		412			91		
$(\Sigma a)^2/n$	0		1,6			0,1		
σ^2	1,55		45,6			10,1		

$= 600/10 = 60$; $576/10 = 57,6$; $569/10 = 56,9$. Средние округляем, получаем величины A : 60; 58; 57, которые вычитаем из величин соответствующего ряда $a = x - A$, далее находим квадраты отклонений и суммы Σa , Σa^2 . Дисперсии вычисляем по формуле (§ 1.16):

$$\sigma^2 = \frac{\Sigma a^2 - (\Sigma a)^2 / n}{n - 1}.$$

Натуральные логарифмы дисперсий находим по таблицам, их сумма равна:

$$\begin{aligned} \Sigma \ln \sigma^2 &= \ln 1,55 + \ln 45,6 + \ln 10,1 = \ln 1,55 + \ln (0,456 \cdot 10^2) + \\ &+ \ln (0,101 \cdot 10^2) = 0,43825 - 0,78526 + 4,60517 + 4,60517 - \\ &- 2,29263 = 6,5707. \end{aligned}$$

Суммы квадратов отклонений в данном случае равны для каждого ряда

$$\Sigma (x - M)^2 = \Sigma a^2 - \frac{(\Sigma a)^2}{n},$$

т. е. величине числителя в предыдущей формуле. Отдельные суммы равны 14; 410,4; 90,9 и их сумма $\Sigma \Sigma (x - M)^2 = 515,3$. Число групп в данном случае $g = 3$, число наблюдений в каждой группе $n = 10$. По формулам для сравнения невзвешенных равночисленных

рядов вычислим критерий Бартлетта:

$$\begin{aligned}
 a &= 1 + \frac{1}{3(g-1)(n-1)} \left(g - \frac{1}{g} \right), \\
 b &= \left[g \left(\ln \frac{\sum \sum (x-M)^2}{g(n-1)} \right) - \sum \ln \sigma^2 \right] (n-1), \\
 a &= 1 + \frac{1}{3 \cdot 2 \cdot 9} \left(3 - \frac{1}{3} \right) = 1,04938; \\
 b &= \left[3 \left(\ln \frac{515,3}{3 \cdot 9} \right) - 6,5707 \right] (10 - 1) = 20,484; \\
 b/a &= \frac{20,484}{1,04938} = 19,52.
 \end{aligned} \tag{4.50}$$

Число степеней свободы в данном случае также равно $\nu = g - 1$. По табл. 10П находим для $\nu = 2$ и уровня значимости 0,05 табличную величину $\chi_{0,05}^2(2) = 5,99$, что меньше вычисленной. Следовательно, условия года оказывают существенное влияние на величину соцветий гелениума.

Если потребуется вычислить критерий Бартлетта для сравнения невзвешенных рядов разной численности, то следует применить формулу (4.49), в которой величины $f = 1$, т. е. частоты будут отсутствовать в расчетах.

§ 4.21. Ранговый критерий Фридмана

Рассматриваемый критерий применяется для одновременного сравнения некоторого числа групп между собой, чтобы выявить степень однородности данного комплекса по варьированию составляющих его данных. Число сравниваемых групп, входящих в изучаемый комплекс, может быть любым, но желательно, чтобы оно было больше 4 и не очень велико, так как большое число групп и данных в них затрудняет ранжировку. Число данных во всех группах должно быть одинаковым. При большом числе групп и данных в них лучше применить критерий Бартлетта (§ 4.20) или дисперсионный анализ. Критерий Фридмана дает менее точные результаты, чем названные методы, однако он свободен от ограничений, связанных с типом распределения исходных данных. В основном критерий Фридмана применяют для получения предварительных выводов, на которые нецелесообразно тратить много времени. В большинстве биологических исследований, тем не менее, точность выводов, оцениваемых при помощи критерия Фридмана бывает обычно достаточной.

Технику расчетов рассмотрим на примере из фенологии видов лиственницы, произраставших в 60-х годах текущего столетия в дендрариуме Ботанического института АН СССР в Ленинграде. В первой части табл. 4.30 приведены (в днях от 1 марта) даты начала роста побегов у 6 видов лиственницы за 7 лет.

1. Выявим степень однородности этих фенодат сначала по годам наблюдений, для чего даты ранжируем (во второй части таблицы) по

Таблица 4.30. Анализ фенологии начала роста побегов у лиственницы с помощью критерия Фридмана

Вид лиственницы	Даты, в днях от 1 марта, по годам							
	1954 г.	1955 г.	1956 г.	1957 г.	1958 г.	1959 г.	1960 г.	
Л. Чекановского	67	74	68	65	75	49	57	
Л. европейская	69	77	69	67	71	49	56	
Л. даурская	69	78	75	67	76	60	64	
Л. американская	71	78	75	71	76	60	70	
Л. японская	71	81	78	70	80	62	68	
Л. сибирская	66	78	75	67	76	52	60	
Баллы по строкам (влияние годовых различий погоды) $n = 6; g = 7$	4 4,5 4 3,5 4 3	6 7 7 7 7 7	5 4,5 5 5 5 5	3 3 3 3,5 3 4	7 6 6 6 6 6	1 1 1 1 1 1	2 2 2 2 2 2	
Суммы R	23	41	29,5	19,5	37	6	18	R
Баллы по столбцам (влияние видовых различий) $n = 7; g = 6$	2 3,5 3,5 5,5 5,5 1	1 2 4 4 6 4	1 2 4 4 6 4	1 3 3 6 5 3	2 1 4 4 6 4	1,5 1,5 4,5 4,5 6 3	2 1 4 6 5 3	10,5 14 27 34 39,5 22

строкам, от меньшего к большему числу. Так, у лиственницы Чекановского рост побегов начался раньше в 1959 г., которому присваиваем ранг 1, затем в 1960 г., которому соответствует ранг 2, затем в 1957 г. — ранг 3, и т. д. по всем строкам. По каждому виду ранжирование проводится отдельно, в пределах одной строки. Если даты совпадают, как, например, у лиственницы европейской в 1954 и 1956 гг., то присваиваем им один и тот же средний ранг: $(4 + 5)/2 = 4,5$. Суммируем ранги по столбцам, получаем суммы: за 1954 г. $R = 23$, за 1955 г. $R = 41$ и т. д. Определяем число групп, степень различия между которыми нас интересует. В данном случае это число лет наблюдений: $g = 7$. Число наблюдений в каждой группе равно $n = 6$. Вычисляем величины хи-квадрат по формуле:

$$\chi^2 = 3 \left[\frac{4 \sum R^2}{ng(g+1)} - n(g+1) \right], \quad (4.51)$$

$$\sum R^2 = 23^2 + 41^2 + 29,5^2 + 19,5^2 + 37^2 + 6^2 + 18^2 = 5189,5;$$

$$\chi^2 = 3 \left[\frac{4 \cdot 5189,5}{6 \cdot 7(7+1)} - 6(7+1) \right] = 41,34.$$

Число степеней свободы, необходимое для оценки величины хи-квадрат по табл. 10П или 10Па определяется по формуле $\nu = g - 1$,

т. е. в нашем случае $\nu = 7 - 1 = 6$. При 0,05 уровне значимости и 6 степенях свободы табличное значение хи-квадрат равно 12,592, что значительно меньше вычисленной величины 41,34. Следовательно годовые различия погоды оказывают достоверное влияние на даты начала роста побегов у видов лиственницы в Ленинграде. Если вычисленная величина хи-квадрат будет меньше его табличного значения, то действие данного фактора следует считать невыявленным.

2. Сроки наступления фенодат, конечно, различаются также и в зависимости от генотипических, видовых особенностей. Чтобы выявить степень однородности данного массива фенодат в отношении видовых различий, проведем в третьей части табл. 4.30 ранжирование данных по столбцам. Так, в 1954 г. раньше чем у других видов начался рост побегов у лиственницы сибирской, которой присвоен 1 балл, затем последовали лиственница Чекановского (балл 2), европейская и даурская (средний балл $= (3 + 4)/2 = 3,5$), далее американская и японская (средний балл $= (5 + 6)/2 = 5,5$). Подобно этому ранжированы даты по другим годам. Суммы рангов по строкам приведены в последнем столбце третьей части таблицы; сумма их квадратов равна:

$$\Sigma R^2 = 10,5^2 + 14^2 + 27^2 + 34^2 + 39,5^2 + 22^2 = 4235,5.$$

Число групп (видов) теперь равно $g = 6$, а число наблюдений $n = 7$. По формуле (4.51):

$$\chi^2 = 3 \left[\frac{4 \cdot 4235,5}{7 \cdot 6 (6 + 1)} - 7 (6 + 1) \right] = 25,88.$$

Число степеней свободы $\nu = (g - 1) = 6 - 1 = 5$. По таблице 10П для уровня значимости 0,05 находим табличное значение хи-квадрат, равное 11,1, что меньше эмпирической величины 25,88. Следовательно, на сроки наступления фенофазы — начало роста побегов достоверное влияние оказывают также и видовые различия, в пределах данной группы из 6 видов лиственницы.

Возникает вопрос, какой из двух рассмотренных факторов влияет сильнее на сроки начала роста побегов. Приблизительное представление об относительном влиянии обоих факторов можно составить при помощи коэффициента (Γ) относительной силы действия факторов:

$$\Gamma = \frac{\chi_{\text{э}}^2}{\chi_{\text{т}}^2},$$

где $\chi_{\text{э}}^2$, $\chi_{\text{т}}^2$ вычисленная и табличная величины хи-квадрата. Для 1 фактора (различия по годам): $\Gamma = 41,3/12,6 = 3,3$; для 2 фактора (видовые различия): $\Gamma = 25,9/11,1 = 2,5$. Следовательно, на сроки начала роста побегов у лиственницы в Ленинграде достоверное влияние оказывают как годовые различия погоды, так и видовые особенности, однако несколько сильнее действуют годовые различия погоды.

Заметим, что при небольшом числе групп ($g \leq 4$) достоверность полученной по методу Фридмана величины хи-квадрат можно оце-

нивать при помощи более точных, чем 10П и 10Па, специальных таблиц, которые приведены, например, в пособии П. В. Терентьева и Н. С. Ростовской (1977) и справочнике А. Бернштейна (1968).

§ 4.22. Ранговый критерий Крускала—Уоллиса

Ранговый критерий применяется для одновременного сравнения некоторого числа групп данных, распределение которых должно соответствовать нормальному типу. Число сравниваемых групп может быть любым, но желательно, чтобы оно было не меньше 4. Численность групп при этом может быть различной, но желательно, чтобы объем меньшей из них был не меньше 5. При большом числе сравниваемых групп и данных в них работа по ранжированию чисел становится затруднительной и может оказаться более эффективным применение критерия Бартлетта. Следует также учитывать меньшую точность результатов сравнения по всем вообще ранговым критериям. Затрудняет расчеты по критерию Крускала—Уоллиса также большое число совпадающих данных, которым надо присвоить одинаковый средний ранг. В общем этот критерий применяют в тех же случаях, что и критерий Фридмана, но при разной численности сравниваемых групп. Рассмотрим процедуру применения критерия Крускала—Уоллиса для сравнения измерений за 4 года диаметра соцветий (в мм) у гелениума осеннего, сорт Ротхаут.

В табл. 4.31 приведены эти данные (столбцы 1—4) и соответствующие им ранги (столбцы 5—8). Ранжирование ведется от меньшего к большему числу, все числа при этом принимаются за единый массив, без деления на группы. Совпадающим данным присваиваем средний ранг: $(2+3)/2 = 2,5$; $(4+5)/2 = 4,5$; $(7+8+9+10)/4 = 8,5$; $(12+13+14+15+16)/5 = 14$; $(17+18+19)/3 = 18$; $(20+21+22+23+24+25+26+27+28+29+30+31+32)/13 = 39,5$; $(33+34+35+36)/4 = 34,5$; $(37+38)/2 = 37,5$; $(39+40)/2 = 39,5$; $(42+43)/2 = 42,5$; $(44+45)/2 = 44,5$; всего получилось 11 групп объединенных рангов. Находим по каждому году сумму рангов W . Общий объем данного массива $N = \sum n = 10+9+11+15 = 45$. Вычисляем величину хи-квадрат:

$$\chi^2 = 3 \left[\frac{4 \sum \frac{W^2}{n}}{N(N+1)} - N - 1 \right], \quad (4.52)$$

$$\sum \frac{W^2}{n} = \frac{264,5^2}{10} + \frac{168,5^2}{9} + \frac{379^2}{11} + \frac{223^2}{15} = 26524,259;$$

$$\chi^2 = 3 \left[\frac{4 \cdot 26524,259}{45(45+1)} - 45 - 1 \right] = 15,7638.$$

В том случае, когда при ранжировании окажутся группы объединенных рангов, которых у нас 11, требуется разделить полученную величину хи-квадрат на поправку:

$$1 - \frac{\sum T}{N^3 - N}, \quad (4.53)$$

Таблица 4.31. Сравнение диаметра соцветий у гелениума за 4 года посредством критерия Крускала—Уоллиса

Годы				Ранги			
1	2	3		1	2	3	4
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
62	54	65	55	37,5	4,5	42,5	6
58	59	61	58	14	18	34,5	14
60	60	56	56	26	26	8,5	8,5
60	44	66	60	26	1	44,5	26
60	65	62	60	26	42,5	37,5	26
61	50	66	59	34,5	2,5	44,5	18
60	63	63	59	26	39,5	39,5	18
61	60	60	54	34,5	26	26	4,5
60	56	64	50	26	8,5	41	2,5
58		61	58	14		34,5	14
		60	60			26	26
			57				11
			56				8,5
			60				26
			58				14
$n = 10$	$n = 9$	$n = 11$	$n = 15$	$W = 264,5$	$W = 168,5$	$W = 379$	$W = 223$

где $T = (t - 1) t (t + 1)$, t — число совпадающих рангов в каждой группе объединенных рангов. Для каждой группы объединенных рангов, в которой совпало 2 ранга, $T = (2 - 1) 2 (2 + 1) = 6$; таких групп у нас 6. При совпадении 3 рангов $T = (3 - 1) 3 (3 + 1) = 24$ таких групп у нас одна, при совпадении 4 рангов $T = (4 - 1) 4 (4 + 1) = 60$ таких групп 2, при 5 совпадениях $T = (5 - 1) 5 (5 + 1) = 120$ таких групп одна и в одной группе совпадают 13 рангов: $T = (13 - 1) \times 13 (13 + 1) = 2184$. Таким образом: $\Sigma T = 6 \cdot 6 + 24 + 60 \cdot 2 + 120 + 2184 = 2484$.

По формуле (4.53):

$$1 - \frac{2484}{45^3 - 45} = 0,9727.$$

Делим величину хи-квадрат на полученную поправку:

$$\frac{15,7638}{0,9727} = 16,2.$$

Последнюю, уточненную величину хи-квадрат при числе степеней свободы: $\nu = g - 1 = 4 - 1 = 3$ (где g — число сравниваемых групп) оцениваем по таблице 10П, где на уровне значимости 0,05 критическая величина равна 7,815, что меньше вычисленной 16,2. Следовательно диаметр соцветий гелениума достоверно различается в зависимости от условий года, в котором выращивали растения.

Существуют таблицы, при помощи которых можно оценить величину хи-квадрат по методу Крускала—Уоллиса для $g = 3$ и $n \leq 5$, однако при малом числе групп более надежные результаты сравнения можно получить посредством критерия Стьюдента (§4.01).

СПЕЦИАЛЬНЫЕ БИОЛОГО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ

В главе излагаются различные сведения, приемы и вспомогательные математические методы, которые могут найти применение в полевых или камеральных условиях исследовательской работы биологов.

§ 5.01. Построение шкал балльной оценки

а. Непараметрические структуры шкал.

В биологии и многих других областях научной и производственной деятельности для качественной оценки состояния объектов нередко используют различные шкалы, которые представляют собой последовательный ряд баллов или аналогичных им словесных выражений. Примерами могут служить шкалы зимостойкости деревьев и кустарников, оценки повреждаемости растений вредителями и болезнями, качества различных сельскохозяйственных и промышленных изделий, антропометрических, медицинских признаков, шкала оценки школьных знаний и многие другие. Разработка каждой отдельной шкалы вызывается конкретными потребностями, поэтому, а также из-за различия соответствующих методических принципов наблюдается большое разнообразие в способах балльной оценки объектов.

Вначале рассмотрим непараметрические структуры шкал, для построения большинства из которых не требуется вычисления обычных параметров совокупности: средней, сигмы, медианы, моды и др. При построении шкал балльной оценки рекомендуется принимать во внимание следующие положения.

Направление возрастания последовательности баллов шкалы должно логически соответствовать ее названию, как это, например, имеет место в шкале оценки знаний, где меньшим и большим знаниям соответствуют меньший (1) и больший (5) баллы. В то же время встречаются шкалы, в частности для оценки зимостойкости растений, в которых наибольшей зимостойкости соответствует наименьший балл и наоборот. Однако такие шкалы более правильно было бы называть шкалами незимостойкости.

Баллы шкалы желательно отсчитывать от нуля, что удобнее при математической обработке и графическом представлении материала. Например, применяемая обычно пятибалльная шкала оценки знаний выглядела бы тогда как: 0, 1, 2, 3, 4 (вместо 1, 2, 3, 4, 5).

За число баллов в шкале можно принять близкое к этому понятие числа классов в статистической совокупности. Число классов

взвешенного вариационного ряда (о его составлении см. § 1.01) в основном находят следующими способами.

1. Эмпирическое правило, выработанное на основании большого опыта, по которому число классов следует принимать в пределах $k = 12 \pm 3$.

2. Согласно правилу Стургеса, объем выборки (N) может быть выражен через число классов по формуле $N = 2^{k-1}$, откуда $k = 3,32 \lg N + 1$ или $k = 1,44 \ln N + 1$, что применялось в формуле (1.02).

3. Число классов принимают также равным: $k = \sqrt{N}$ (Дюерфель, 1969), что использовано при построении формулы (1.03) и $k = 5 \lg N$ (Брукс и Карузерс; цит. по: Шторм, 1970).

4. При известных параметрах ряда рекомендуется формула классового интервала:

$$c = 0,6 \sigma, \quad (5.01)$$

которая предусматривает разделение выборки в среднем на 10 классов, где c — величина классового интервала; σ — среднее квадратическое отклонение.

Основным критерием правильности выбора числа классов и, следовательно, оптимального числа баллов является верная передача типа распределения частот данной совокупности: без излишних мешающих подробностей, что является следствием слишком большого числа классов, и без чрезмерного упрощения, что бывает при недостаточном числе классов. Оптимальное число классов в распределениях биологических признаков наиболее часто бывает 10—12; при наименьшем числе классов — 5 и наибольшем — 21 класс (по опыту работы с ботаническими объектами).

Интервалы вариационного ряда делаются обычно равными. Однако баллы большинства шкал имеют неравномерное количественное содержание, которое увеличивается с каждым баллом в сторону усиления признака, или же сначала интервал увеличивается, а затем симметрично уменьшается. В связи с этим возможно применение двух следующих структур шкал.

1. Хотя неравномерное количественное содержание баллов почти всегда устанавливали, не руководствуясь каким-либо теоретическим обоснованием, в этом факте, вероятно, находит свое применение психофизический закон Фехнера, согласно которому сила ощущения каких-либо раздражителей изменяется в арифметической прогрессии, в то время как сила явления, вызывающего раздражение, изменяется в том же направлении в геометрической прогрессии. Используя закон Фехнера, количественное содержание баллов шкалы с достаточным для практики приближением можно рассчитать по правилу геометрической прогрессии, которая задается последним членом n , начальным членом a , числом членов k и знаменателем прогрессии d :

$$\lg d = \frac{\lg n - \lg a}{k - 1} \quad (5.02)$$

Далее будет приведен пример расчета шкалы со структурой геометрической прогрессии.

2. В качестве другой структуры шкалы, когда интервалы между границами баллов сначала увеличиваются, а затем симметрично уменьшаются, применяют шкалу Кристля (Сепетлиев, 1968). В этом случае выборка дат изучаемого признака делится на группы, имеющие следующее количественное содержание (в процентах от всего объема выборки): 3, 7, 15, 50, 15, 7, 3 или 3, 7, 15, 50, 85, 93, 97, 100. Последний ряд указывает границы шкалы, которая является, таким образом, восьмибалльной. Далее будет приведен пример расчета шкалы Кристля.

Для определения границ баллов шкалы и выбора начала ее отсчета важное значение имеет понятие порога чувствительности, интервала безразличия, или предела ощущаемости, которое в терминах математической статистики можно считать уровнем значимости. В дополнение к трем стандартным уровням значимости $W_1 = 5\%$; $W_2 = 1\%$; $W_3 = 0,1\%$ для построения шкал целесообразно ввести еще один уровень значимости — $W_0 = 10\%$, при котором за интервал безразличия, или допускаемую ошибку, принимается 10% ; это также та меньшая разница в величинах, которую в среднем еще надежно улавливают при глазомерных наблюдениях биологических объектов. Более точно интервал безразличия в каждом конкретном случае не трудно определить специальными опытами.

Распределение вариантов изучаемой совокупности по процентным группам в соответствии с выбранной структурой шкалы удобно производить способом процентилей по формуле (1.180):

$$x_p = \frac{(i_p - f_n)c}{f} + g,$$

где x_p — значение процентильной варианты; g — нижняя (меньшая или левая) граница процентильного класса; f_n — сумма частот до процентильного класса; c — классовый интервал; i_p — порядковый номер процентильной варианты; f — частота процентильного класса (не накопленная).

В качестве примера рассчитаем шкалу оценки величины соцветий у нивяника для сортооценки этого декоративного растения.

1. В табл. 5.01 приводится ряд распределения 500 соцветий по их диаметру. Число классов было найдено по формуле (1.02).

2. За основу построения шкалы сначала примем структуру Кристля с изменением: вместо 7 и 15% возьмем соответственно 10 и 25%, так что процентное распределение вариантов по градациям шкалы будет таким: 3, 10, 25, 50, 25, 10, 3, или нарастающим итогом: 3, 10, 25, 50, 75, 90, 97.

Найдем номера процентильных вариантов x_i , где i определяется из выражения:

$$i = \frac{NP^0}{100}, \quad (5.03)$$

где i — номер процентильной варианты; N — объем выборки; P^0 — процентное выражение процентиля.

Таблица 5.01. К вычислению процентилей

Диаметр сопветий, мм	Частоты, f	Ряд накопленных частот, $f\Sigma$	Положение процентильных вариант, x_i
1	2	3	4
30—43	2	2	—
44—57	29	31	x_{15}
58—71	119	150	x_{50}, x_{125}
72—85	171	321	x_{250}
86—99	121	442	x_{375}
100—113	42	484	x_{450}
114—128	16	500	x_{485}
$c=14$	$N=500$		

Для данного случая: $i = \frac{500}{100} \cdot P^0 = 5P^0$, откуда $i_3 = 5 \cdot 3 = 15$; $i_{10} = 5 \cdot 10 = 50$; $i_{25} = 5 \cdot 25 = 125$; $i_{50} = 5 \cdot 50 = 250$; $i_{75} = 5 \cdot 75 = 375$; $i_{90} = 5 \cdot 90 = 450$; $i_{97} = 5 \cdot 97 = 485$.

Номера процентильных вариант можно найти также по несколько более точной формуле (1.179).

3. Для каждого номера процентильной варианты отыщем класс, в котором она должна находиться. Удобно это делать с помощью ряда накопленных частот, который составлен в столбце 3 табл. 5.01. Вариант x_{15} находится в классе 44—57, x_{50} — в классе 58—71 и т. д., как указано в столбце 4. Распределение процентильных вариант по классам нужно, чтобы определить начало процентильного класса g и прочие данные для расчета по формуле (1.180).

4. Расчет процентилей в фактических числах ведется следующим образом: для $P^0 = 3\%$ процентильная варианта x_{15} находится в классе 44—57, следовательно, $g = 44$; $i = 15$; $f_n = 2$; $c = 14$; $f = 29$; откуда по формуле (1.180)

$$x_3 = \frac{(15 - 2) \cdot 14}{29} + 44 = 50,276;$$

$$x_{10} = \frac{(50 - 31) \cdot 14}{119} + 58 = 60,235;$$

$$x_{25} = \frac{(125 - 31) \cdot 14}{119} + 58 = 69,059;$$

$$x_{50} = \frac{(250 - 150) \cdot 14}{171} + 72 = 80,187;$$

$$x_{75} = \frac{(375 - 321) \cdot 14}{121} + 86 = 92,248;$$

$$x_{90} = \frac{(450 - 442) \cdot 14}{42} + 100 = 102,67;$$

$$x_{97} = \frac{(485 - 484) \cdot 14}{16} + 114 = 114,88.$$

5. Округляя полученные процентиля до практически применяемой точности, получим искомую шкалу (табл. 5.02). Если окажется, что рассчитанная шкала слишком детальна, то ее можно упростить

Таблица 5.02. Шкала сортооценки
нивяника по диаметру соцветий,
составленная по процентиям
Кристля

Балл	Диаметр соцветий, мм,	Процент соцветий с данным диамет- ром в общей со- вокупности
1	0—49	3
2	50—59	7
3	60—69	15
4	70—79	25
5	80—91	25
6	92—101	15
7	102—114	7
8	115 и выше	"

Таблица 5.03. Шкала оценки
размера соцветий нивяника,
составленная по геометрической
прогрессии процентилей

Балл	Диаметр соцветий, мм	Процент соцветий с данным диамет- ром в о щей со- вокупности
1	0—48	2
2	49—58	4
3	59—64	11
4	65—80	33
5	81—97	33
6	98—109	11
7	110—119	4
8	120 и выше	2

за счет попарного объединения баллов. Вместо структуры Кристля, когда необходимо рассчитать шкалы с иным числом баллов, возможно применение следующих рядов процентилей: 10, 25, 50, 75, 90 (6-балльная шкала); 1, 3, 10, 25, 50, 75, 90, 97, 99 (10-балльная шкала); 1, 3, 10, 15, 25, 50, 75, 85, 90, 97, 99 (12-балльная шкала).

Как ранее указывалось, при определении числа баллов шкалы лучше руководствоваться формулой (1.02), по которой находят число классов взвешенных рядов вообще, а при назначении конкретных границ баллов можно использовать принцип геометрической прогрессии. Принимая за основу два этих принципа, вместо структуры Кристля рассчитаем другую шкалу сортооценки величины соцветий нивяника.

1. Число баллов определялось при составлении вариационного ряда по формуле (1.02).

2. При помощи геометрической прогрессии (5.02) надлежит построить ряд процентилей, который определит границы баллов. Значение n берется в процентах в зависимости от формы кривой распределения. Если кривая распределения двускатная, то $n = 50\%$, так как рассчитывается лишь половина процентилей шкалы, а вторая половина процентилей ряда принимается симметрично первой, причем k следует взять равным ближайшему четному числу. Если кривая распределения односкатная, то процентили можно рассчитывать сразу на всю шкалу при $n = 100\%$ и без изменения k до ближайшего четного числа.

Рассчитаем две шкалы: сначала для ряда с двускатной кривой распределения частот (диаметр соцветий нивяника), а затем для ряда с односкатной кривой распределения (число детки на одну луковицу у гладнолуca).

1. Число классов ряда распределения диаметров соцветий нивяника, определенное ранее по формуле (1.02), равно $k = 7$. Примем $k = 8$ как ближайшему четному числу, следовательно, число баллов в искомой шкале будет также равно 8.

2. Интервал безразличия, или первый член геометрической прогрессии, возьмем $a = 2\%$.

3. Последний член прогрессии $n = 50\%$, так как рассчитываем шкалу по ряду с двускатной симметричной кривой распределения частот. Следовательно, шкала рассчитывается для половины ее баллов, поэтому $0,5 k = 0,5 \cdot 8 = 4$. Таким образом, все исходные данные получены и знаменатель прогрессии

$$\lg d = \frac{\lg 50 - \lg 2}{4 - 1} = 0,466; \quad d = 2,924.$$

Члены этой геометрической прогрессии получают последовательным умножением их на знаменатель d : 2; 5,848; 17,099; 49,997. Округляя полученные числа и распространяя их симметрично на другую половину ряда, получим искомый ряд процентилей: 2, 6, 17, 50, 83, 94, 98%, которые образуют структуру восьмибалльной шкалы.

4. По формуле (5.03), где $N/100 = 500/100 = 5$; находим номера процентильных вариантов: $i_2 = 5 \cdot 2 = 10$; $i_6 = 5 \cdot 6 = 30$; $i_{17} = 5 \cdot 17 = 85$; $i_{50} = 5 \cdot 50 = 250$; $i_{83} = 5 \cdot 83 = 415$; $i_{94} = 470$; $i_{98} = 490$.

5. Процентильная варианта x_{10} находится в классе 44—57 (табл. 5.01), где $g = 44$; $i = 10$; $f_{11} = 2$; $c = 14$; $f = 29$, откуда по формуле (1.180)

$$x_2 = \frac{(10 - 2) \cdot 14}{29} + 44 = 47,9.$$

Точно таким же образом вычисляются все остальные проценти-ли: $x_6 = 57,5$; $x_{17} = 64,4$; $x_{50} = 80,2$; $x_{83} = 96,9$; $x_{94} = 109,3$; $x_{98} = 119,3$, которые составляют границы баллов шкалы, приводимой в табл. 5.03.

Как видим, шкала, составленная по структуре геометрической прогрессии, более точно оценивает крайние отклонения диаметра соцветий (первый и восьмой баллы охватывают по 2% совокупности) и менее детальна в оценке обычно встречающихся диаметров (центральные баллы по 33%), что можно считать преимуществом этой шкалы по сравнению со шкалой, построенной в табл. 5.02.

В качестве примера рассчитаем шкалу по ряду распределения с односкатной кривой распределения частот с применением структуры геометрической прогрессии процентилей (данные табл. 5.04).

1. Число классов распределения количества детки на луковицу было установлено формулой (1.02): $k = 10$.

2. Для расчета шкал по рядам с односкатной кривой распределения частот, как указывалось выше, применяется формула геометрической прогрессии (5.02), в которой всегда $n = 100\%$, а шкала рассчитывается сразу для всего ряда в противоположность рядам двускатной кривой распределения частот. В нашем примере $n = 100\%$, интервал безразличия возьмем $a = 10\%$, $k = 10$, следовательно:

$$\lg d = (\lg 100 - \lg 10)/(10 - 1) = 0,111; \quad d = 1,291.$$

Таблица 5.04. К расчету шкалы сортов гладиолуса по числу детки на луковицу

Класс чис- ла детки	Середина класса	Частота	Накопленная частота	Положение процентильных вариантов, x_i
0—9,9	5	193	193	49,6; 64,5; 84,3; 109,1; 138,9; 178,6
10—19,9	15	141	334	228,2; 297,6
20—29,9	25	73	407	381,9
30—39,9	35	33	440	—
40—49,9	45	26	466	—
50—59,9	55	11	477	—
60—69,9	65	8	485	—
70—79,9	75	4	489	—
80—89,9	85	3	492	—
90—100	95	4	496	496,0
$k = 10 \quad c = 10 \quad N = 496$				

3. Умножая первый член a и все последующие на знаменатель d , получим ряд процентилей (с округлением до целых чисел): 10, 13, 17, 22, 28, 36, 46, 60, 77, 100%.

4. Номера процентильных вариантов по формуле (5.03) следующие:

$$i_{10} = \frac{496}{100} \cdot 10 = 4,96 \cdot 10 = 49,60; \quad i_{13} = 64,48;$$

$$i_{17} = 84,32; \quad i_{22} = 109,12; \quad i_{28} = 138,88; \quad i_{36} = 178,56;$$

$$i_{46} = 228,16; \quad i_{60} = 297,60; \quad i_{77} = 381,92; \quad i_{100} = 496,00.$$

В столбце 5 табл. 5.04 указано, в каких классах находятся процентильные варианты, номера которых в таблице даются округленно.

5. Процентиль в фактических числах для i_{10} по формуле (1.180):

$$x_{10} = \frac{(49,60 - 0) \cdot 10}{193} + 0 = 2,57.$$

Вычисляя остальные процентиля таким же образом, получим:

$x_{13} = 3,34; \quad x_{17} = 4,37; \quad x_{22} = 5,65; \quad x_{28} = 7,20; \quad x_{36} = 9,25;$
 $x_{46} = 12,49; \quad x_{60} = 17,42; \quad x_{77} = 26,56; \quad x_{100} = 100,00.$ Округ-
 ляя полученные границы баллов, составим шкалу, приведенную
 в табл. 5.05.

Таблица 5.05. Шкала оценки сортов гладиолуса по количеству детки на луковицу

Балл	Количество детки на луковицу	Растения с данным количеством детки на одну луковицу, %	Балл	Количество детки на луковицу	Растения с данным количеством детки на одну луковицу, %
1	0—2,5	10	6	7,2— 9,2	8
2	2,6—3,2	3	7	9,3— 12,4	10
3	3,3—4,3	4	8	12,5— 17,3	14
4	4,4—5,6	5	9	17,4— 26,5	17
5	5,7—7,1	6	10	26,6—100	27

Таблица 5.06. К вычислению процентилей по частотам нормальной кривой

Класс, диаметр соцветий, мм	Частота нормальной кривой	Ряд накопленных частот	Положение процентильных вариантов, x_i
30—43	4,48	4,48	—
44—57	31,57	36,05	x_{10} ,
58—71	106,14	142,19	x_{85}
72—85	170,38	312,57	x_{250}
86—99	130,58	443,15	x_{415}
100—113	47,76	490,91	x_{470} ,
114—128	8,34	499,25	—
14	$N = 499,25$		

Если суммировать проценты в столбце 3, то результат будет несколько больше 100%, что вызвано округлением процентов до целого числа. Количество детки в шкале дается в дробных числах, так как оценка делается на основании средних величин.

Все вышерассмотренные шкалы были рассчитаны по эмпирическим или фактическим частотам рядов распределений. Такие частоты нередко содержат случайные отклонения от общей тенденции их распределения. Для того чтобы по возможности освободить шкалу от влияния случайных отклонений, ее следует рассчитать по теоретической кривой распределения.

Рассчитаем шкалу, построенную на основании теоретических частот нормального распределения соцветий пивняка по диаметру.

1. Определение числа классов ряда, приведенного в табл. 5.06, сделано по формуле (1.02); принимаем $k = 8$ с округлением до четного числа.

2. В качестве структурной основы шкалы возьмем проценты, рассчитанные выше для этого же ряда по геометрической прогрессии: 2, 6, 17, 50, 83, 94, 98%.

3. Номера процентильных вариантов по формуле (5.03), также вычисленные выше (округленно): $i_2 = 10$; $i_6 = 30$; $i_{17} = 85$; $i_{50} = 250$; $i_{83} = 415$; $i_{94} = 470$; $i_{98} = 490$.

Положение процентильных вариантов в ряду указано в столбце 4 табл. 5.06.

4. По формуле (1.180) были получены следующие проценты: 46,448; 55,317; 64,457; 80,859; 96,982; 107,871; 114,026. Округляя эти числа, составим шкалу (табл. 5.07).

В заключение кратко перечислим способы, которыми могут быть рассчитаны шкалы.

1. На основе структуры Крестля: 3, 7, 15, 50, 15, 7, 3% или аналогичного ряда процентилей с произвольным их числом, следовательно, и числом баллов.

2. На основе ряда процентилей, образованного геометрической прогрессией, начиная от величины интервала безразличия, или уровня значимости. Число баллов задается формулой (1.02):

Таблица 5.07. Шкала оценки соцветий нивяника по диаметру, составленная по частотам нормальной кривой

Балл	Диаметр соцветий, мм	Соцветия с данным диаметром в общей совокупности, %	Балл	Диаметр соцветий, мм	Соцветия с данным диаметром в общей совокупности, %
1	0—46	2	5	81— 96	33
2	47—55	4	6	97—107	11
3	56—64	11	7	108—114	4
4	65—80	33	8	114 и выше	2

а) для рядов с двускатной кривой распределения частот шкала рассчитывается для половины ряда, вторая половина образуется симметрично первой, число баллов (если оно не четное) округляется до ближайшего четного числа;

б) для рядов с односкатной кривой распределения частот шкала рассчитывается сразу по всему ряду, число классов не изменяется.

3. Каждый из способов (1, 2а, 2б) может быть применен двояко — к эмпирическим или к теоретическим частотам данного ряда. Если кривая эмпирических частот имеет неровности и заметные случайные отклонения, то ее рекомендуется выровнять любой подходящей кривой распределения или даже какой-либо другой кривой и шкалу рассчитать по теоретическим частотам.

б) Параметрические структуры шкал. Сигмальные монотипичные шкалы.

В отличие от всех структур шкал, рассмотренных выше в данном параграфе, сигмальная структура шкал наиболее теоретически обоснована, так как исходит из свойств нормального распределения и общего свойства всех кривых, обладающих точками перегиба. Это свойство, или правило, заключается в том, что точки перегиба кривых порядка второго и выше отражают границы качественного изменения процесса, который аппроксимируют данные кривые. Кривая нормального распределения, в основном моделирующая процесс изменения величины отклонений вариант от средней, отражает также и те моменты, когда варианты настолько малы или настолько велики, что начинают становиться нетипичными для данной совокупности, хотя и остаются еще принадлежащими к ней. Эти моменты графически и аналитически отражаются на ходе кривой точками перегиба: $-\sigma$, $+\sigma$, единственными на каждой из ветвей симметричной нормальной кривой (см. рис. 4).

В рассматриваемом здесь аспекте необходимо также введение понятия типичности вариант, дат или любых объектов, из которых можно составлять статистические совокупности. При этом центру типичности в терминах математической статистики будет соответствовать средняя данной совокупности, или ее математическое ожидание. Основными границами типичности и нормы в обе стороны от средней является величина среднего квадратического отклонения

$(\pm \sigma)$, а на графике — соответствующие точки перегиба нормальной кривой. В соответствии с установленной здесь исходной единицей, или квантом $(\pm \sigma)$ типичности и нормы, на его основе возможно построение шкал с различным числом баллов.

Самая простая — альтернативная, или двухбалльная, шкала, может быть построена с границами типичности: $x_i < M$, $x_i > M$, т. е. варианты классифицируют по признаку: больше или меньше средней. Трехбалльная шкала строится при классификации вариантов на три группы: $x_i < M - \sigma$, $M - \sigma < x_i < M + \sigma$, $x_i > M + \sigma$. Границы четырехбалльной шкалы: $x_i < M - \sigma$, $M - \sigma < x_i < M$, $M < x_i < M + \sigma$, $M + \sigma < x_i$. Оптимальной шкалой, наиболее соответствующей особенностям нормального распределения, вероятно следует считать шестибалльную шкалу с границами $\pm \sigma$, $\pm 2\sigma$, $\pm 3\sigma$ в обе стороны от средней (M). Если учитывать также и участки кривой влево от -3σ и вправо от $+3\sigma$, то указанная шкала станет восьмибалльной. Однако такие объекты оценки, которые бы соответствовали этим двум крайним баллам, могут встретиться очень редко. Шестибалльная шкала схематично представлена на рис. 4. Баллы 3 и 4 совместно охватывают 68,3% всех вариантов совокупности и представляют собой диапазон нормы, или типичности, величин признака по шестибалльной шкале. Дальнейшее увеличение числа баллов шкалы может быть достигнуто уменьшением ее интервала, например, 12-балльная шкала может быть построена с интервалом через $0,5\sigma$; с учетом значений меньше -3σ и больше $+3\sigma$ эта шкала становится 14-балльной.

Не давая теоретического обоснования, но тем не менее правильно применили сигмальные шкалы в антропологии Мартин (Martin, 1928) и Ю. Л. Поморский (1930).

Приведем семибалльную шкалу Мартина для оценки развития человека, которая может найти в ряде случаев и более широкое применение: 1) необычайно хорошее развитие: от ∞ до $M + 2,5\sigma$; 2) очень хорошее: от $M + 2,5\sigma$ до $M + 1,5\sigma$; 3) хорошее: от $M + 1,5\sigma$ до $M + 0,5\sigma$; 4) среднее: от $M + 0,5\sigma$ до $M - 0,5\sigma$; 5) плохое: от $M - 0,5\sigma$ до $M - 1,5\sigma$; 6) очень плохое: от $M - 1,5\sigma$ до $M - 2,5\sigma$; 7) необычно плохое: от $M - 2,5\sigma$ до 0.

Шкала, в основу которой положено симметричное эмпирическое распределение вариантов, обладает свойством также и симметрии типичности относительно ее центрального балла. Чем более симметрична кривая эмпирического распределения, тем ближе к центру шкалы находится средняя арифметическая этого распределения и, следовательно, центр типичности. Напротив, у асимметричных распределений дат центр типичности и его границы $(\pm \sigma)$ смещены в соответствии со степенью асимметрии, но метод построения сигмальных шкал остается при этом таким же, как и у симметричных распределений.

Сигмальные политипичные шкалы. Все биологические объекты, как правило, многопризнаковые, поэтому для их оценки предпочтительнее применять политипичные шкалы. Сигмальные шкалы дают также возможность построить и теоретически обоснованную

политипичную шкалу оценки объектов по многим признакам, в отличие от рассмотренной выше монотипичной шкалы, предназначенной для оценок по одному признаку.

При классификации величин какого-либо признака необходимо предварительно решить, что следует оценивать: типичность или норму признака. Для целей систематики, например, не имеет особого значения знак отклонения величин признаков от их центров типичности (средних арифметических), а при оценке достоинств сортов растений, напротив, важно знать, лучше данный сорт по комплексу признаков или хуже некоторого другого сорта. Комплексную оценку типичности объекта можно вести при помощи показателя атипичности (Φ), вычисляемого по формуле

$$\Phi = \frac{1}{n} \sum_i^n \left| \frac{a_i - M_i}{\sigma_i} \right|, \quad (5.04)$$

где n — число признаков; i — порядковый номер признака; a_i — отдельные значения признака; M_i — средняя арифметическая величин признака; σ_i — среднее квадратическое отклонение величин признака. Величина в прямых скобках должна приниматься всегда положительной. Степень отклонения объекта от нормы по комплексу признаков можно оценить посредством показателя аномальности:

$$\Phi_1 = \frac{1}{n} \sum_i^n \frac{a_i - M_i}{\sigma_i}, \quad (5.05)$$

где обозначения те же, за исключением того, что суммирование нормированных отклонений производится с учетом их знаков.

В литературе даны два примера расчета политипичных шкал на основе показателей Φ и Φ_1 , посредством которых из двух больших совокупностей идентифицированы и отобраны типичные по фенологии травянистые (для Москвы) и древесные (для Ленинграда) виды, и составлена шкала оценки фенологической аномальности травянистых многолетников в Москве (Зайцев, 1978, 1981).

Сигмальные политипичные шкалы в силу своей органичной (благодаря универсальности нормального распределения) связи, в частности с объектами живой природы, могут найти применение для классификации любых объектов, репрезентативные совокупности значений признаков которых выражены количественно, в том числе в различных отраслях ботаники. Если такая классификация связана с принципиальными теоретическими вопросами науки, то возникает необходимость в биологическом обосновании применения метода политипичной сигмальной шкалы.

Так, например, для целей систематики вполне заслуживает испытания один из принципов филогенеза, изложенный В. Л. Комаровым (1908), который считал, что близость к началу процесса обособления данной систематической группы определяется тем, что формы, к ней относящиеся, обладают более общими неопределенными, расплывчатыми признаками, приближающимися к признакам других,

соседних групп. Эта идея вполне может быть положена в основу алгоритма классификации посредством политипичной сигмальной шкалы. Иначе говоря, в понятиях математической статистики, чем ближе вид по комплексу признаков к средней арифметической для данного рода, тем более древним он считается. Следовательно, применяя показатель атипичности Φ , можно попытаться создать естественную классификацию и систему видов внутри рода или другого таксона, а также найти филогенетически наиболее древний родоначальный вид.

Приведем пример применения показателя аномальности для построения шкалы оценки сортов по срокам роста и развития.

Политипичная оценка 28 сортов гелениума осеннего по 5 фенологическим признакам посредством показателя аномальности Φ_1 вычислялась для каждого сорта по формуле:

$$\Phi_1 = \frac{\frac{a - 49,55}{1,844} + \frac{b - 153,39}{5,27} + \frac{c - 206,78}{3,27} + \frac{d - 207,13}{13,95} + \frac{e - 55,74}{5,61}}{5},$$

где a, b, c, d, e — фенологические даты сортов, которые в табл. 5.08 приведены в форме обычных календарных дат. В конце таблицы даны средние арифметические в виде числа дней от 1 марта и сигмы, указанные в формуле (5.05).

Сравнивая показатели аномальности в столбце 8, видим, что наиболее типичен по фенологии сорт Гольдфукс ($\Phi_1 = -0,00339$), а более других отклоняется от нормы сорт Кокарде ($\Phi_1 = 1,09$), фенофазы которого более запаздывают по сравнению с другими сортами. Следовательно, имеются некоторые шансы вывести на его основе новый поздноцветущий сорт, что является актуальным вопросом в селекции сортов гелениума. Самый ранний по комплексу фенофаз — сорт Морхаим Бьюти.

Оценку комплексного показателя аномальности Φ_1 можно производить по 15-балльной шкале, приведенной в табл. 5.09. Это наиболее подробная шкала, которая позволяет детально дифференцировать объекты изучения по сравнительно небольшим их отклонениям от интервала нормы ($M \pm \sigma$), который в обозначениях шкалы из табл. 5.09 определен границами от 1,24 до —1,25. Менее детальная семибалльная шкала приведена в табл. 5.10, где норма определена границами от 1,5 до —1,49 (3, 4, 5 баллы). Оценивая типичность фенологических признаков сортов гелениума осеннего по 15-балльной шкале, видим, что все сорта по величине показателя Φ_1 находятся в пределах нормы, хотя и различаются между собой.

Шкала из табл. 5.09 дает возможность дифференцировать сорта и в пределах нормы. Так как фенофазы многолетников в подавляющем большинстве случаев связаны между собой положительной корреляцией, то можно утверждать, что показатель Φ_1 характеризует в общем раннее и позднее наступление всех фенофаз по сорту. Следовательно, распределив сорта в соответствии с величиной Φ_1 по баллам шкалы (столбец 9 табл. 5.08), получим возможность распределить сорта гелениума по срокам роста и развития: 6 —

Таблица 5.08. Фенология сортов гелениума осеннего

Сорт	Число лет наблюдений *	Начало отрастания	Начало цветения	Конец цветения	Созревание	Продолжительность цветения, дни	Φ , показатель анормальности	Баллы
1	2	3	4	5	6	7	8	9
Альтгольд	21	19.IV	7.VIII	16.IX	23.IX	43	—0,622	7
Альтгольдризе	5	20.IV	24.VII	21.IX	1.X	59	—0,0587	8
Аугустзонне	20	19.IV	5.VIII	24.IX	27.IX	53	0,214	8
Бидермайер	16	18.IV	1.VIII	22.IX	27.IX	52	—0,185	8
Бишоп	2	12.IV	13.VIII	26.IX	30.VIII	63	—0,122	8
Чипперфилд Ориндж	9	22.IV	24.VII	23.IX	20.VIII	76	0,333	9
Ди Блондые	5	21.IV	31.VII	24.IX	27.IX	54	0,337	9
Гартензонне	21	20.IV	1.VIII	19.IX	21.IX	51	—0,299	7
Глётауге	21	19.IV	1.VIII	22.IX	23.IX	54	—0,0694	8
Гольдфукс	21	20.IV	31.VII	23.IX	21.IX	54	—0,00339	8
Гольдфуксфрюе	14	18.IV	25.VII	23.IX	22.IX	57	—0,491	7
Гольдлакцверг	21	19.IV	29.VII	21.IX	23.IX	55	—0,159	8
Гранатштерн	21	18.IV	29.VII	26.IX	22.IX	60	0,0842	8
Гербстрот	21	18.IV	30.VII	27.IX	24.IX	59	0,279	9
Катарина	20	18.IV	31.VII	22.IX	23.IX	54	—0,275	7
Кокарде	14	19.IV	2.VIII	26.IX	22.IX	56	1,09	10
Кугельзонне	5	20.IV	3.VIII	27.IX	26.IX	55	0,528	9
Купфершпрудель	3	18.IV	27.VII	19.IX	10.IX	54	—0,757	7
Лихтгёттин	14	18.IV	31.VII	22.IX	20.IX	57	—0,231	8
Морхайм Бьюти	10	16.IV	25.VII	18.IX	14.IX	57	—0,867	6
Риверстон Бьюти	10	19.IV	31.VII	19.IX	21.IX	54	—0,282	7
Ротхаут	21	20.IV	31.VII	25.IX	24.IX	58	0,327	9
Шпетгольдкуппель	9	20.IV	10.VIII	28.IX	4.X	49	0,750	10
Зептемберзонне	21	19.IV	10.VIII	30.IX	29.IX	54	0,823	10
Зоммерзонне	9	17.IV	30.VII	23.IX	19.IX	58	—0,154	8
Зонненберг	21	18.IV	3.VIII	20.IX	19.IX	51	—0,372	7
Супербе	5	20.IV	27.VII	25.IX	5.X	60	0,465	9
Типика	15	17.IV	3.VIII	20.IX	22.IX	53	—0,283	7

Средняя M	49,55	153,39	206,78	207,13	55,74
Дисперсия σ^2	3,3998	27,788	10,684	194,56	31,458
Среднее квадратиче- ское отклонение σ	1,844	5,27	3,27	13,95	5,61
Коэффициент вариации v	3,72	3,44	1,58	6,74	10,1

* Наблюдения велись в период с 1949 по 1970 г. в отделе цветоводства Главного ботанического сада АН СССР (Москва).

Таблица 5.09. Границы классов распределения нормированного отклонения t и 15-балльная шкала для оценки показателя аномальности Φ_1

Балл	Границы		Сер. дина. класса
	левая	правая	
15	3,75	3,25	3,5
14	3,24	2,75	3
13	2,74	2,25	2,5
12	2,24	1,75	2
11	1,74	1,25	1,5
10	1,24	0,75	1
9	0,74	0,25	0,5
8	0,24	—0,25	0 Норма
7	—0,26	—0,75	—0,5
6	—0,76	—1,25	—1
5	—1,26	—1,75	—1,5
4	—1,76	—2,25	—2
3	—2,26	—2,75	—2,5
2	—2,76	—3,25	—3
1	—3,26	—3,75	—3,5

с фенологией сортов гелениума, применена многобалльная шкала, так как совокупность представляет собой группу сравнительно близких между собой по фенологии растений, в которой не наблюдается больших различий по фенодатам. Это и подтвердилось после распределения Φ_1 по баллам шкалы, так как все сорта вошли в интервал нормы (6 — 10 баллы шкалы в табл. 5.09).

В результате изучения свойств показателей атипичности Φ (без учета знаков отклонений) и аномальности Φ_1 (с учетом знаков отклонения) выявляется также одна существенная особенность их применения в зависимости от изучаемого комплекса данных. Если признаки данного комплекса связаны только отрицательными или только положительными связями, лучше применять показатель Φ_1 . Если же среди взаимосвязей комплекса встречаются как положительные, так и отрицательные связи, причем ни одна из них по

очень ранние, 7 — ранние, 8 — средние, 9 — поздние, 10 — очень поздние, что полностью соответствует реальному ходу сезонного роста и развития этих сортов по совокупности результатов наблюдений основных фенофаз. Из приведенного примера видно, что результаты, более ценные для практики, чаще можно получить, пользуясь более детальной, 15-балльной шкалой. Напротив, менее детальная шкала, т. е. с меньшим числом баллов, может не выявить некоторые важные особенности и различия между единицами совокупности.

В общем случае следует тем более предпочитать многобалльную шкалу, чем меньше дисперсия распределения вариант в данной совокупности. Более варьирующие объекты можно классифицировать по шкале с меньшим числом баллов. В приведенном примере

Таблица 5.10. Границы классов распределения нормированного отклонения и 7-балльная шкала для оценки показателя аномальности Φ_1

Балл	Границы класса		Середина класса
	левая	правая	
7	3,5	2,49	3
6	2,5	1,49	2
5	1,5	0,49	1
4	0,5	—0,49	0 Норма
3	—0,5	—1,49	—1
2	—1,5	—2,49	—2
1	—2,5	—3,5	—3

$$k = 7 \quad c = 1$$

частоте встречаемости явно не преобладает, то лучше вычислять показатель Φ . В случае преобладания в данном комплексе признаков криволинейных двузначных связей также следует вычислять Φ . Для биологических задач все же большую ценность в большинстве случаев представляет собой применение показателя Φ_1 , учитывающего знаки отклонений, так как в биологии часто важно знать, больше или меньше, хуже или лучше нормы данная особь по отдельному признаку или по совокупности последних. Например, о сорте всегда важно знать, каковы сроки его роста и развития по комплексу фенофаз, больше или меньше нормы значения признаков, из которых складывается декоративность сорта, устойчивость к перезимовке, к болезням и многое другое.

Из того факта, что необходимо существует норма в пределах сигмальных отклонений у любого признака, вытекает положение об относительности категории однородности. Абсолютно однородных массивов данных по какому-либо признаку не существует, так как массив реальных данных всегда можно разделить на типичные и нетипичные варианты, поскольку любой признак варьирует и обладает дисперсией своих величин.

§ 5.02. Кодирование признаков жизненных форм высших растений

Математическая обработка информации, получаемой при изучении жизненных форм растений, требующая унификации данных, встречается в настоящее время большие трудности в связи со сложностью и противоречивостью существующих классификаций жизненных форм. Для облегчения этой задачи предлагается упрощенная система обозначений жизненных форм, удобная при любой форме хранения, поиска и обработки информации.

Поскольку понятие «жизненная форма» составное, то его следует отражать также составным символом. Для этого необходимо по возможности четко разграничить признаки, характеризующие жизненную форму, каждый из этих признаков закодировать в баллах логически последовательной шкалы, а затем выражать конкретные жизненные формы составным кодом. Достаточно полно жизненную форму у высших растений могут характеризовать следующие семь признаков.

А. Форма роста

1. Дерево
2. Кустарник
3. Кустарничек
4. Полукустарник и полукустарничек
5. Травянистое

Б. Дополнительные признаки формы роста

1. Прямостоячее
2. Лиана

3. Висячее

4. Стелющееся
5. Розеточное
6. Подушковидное
7. Клубнеобразующее
8. Луковичное

В. Продолжительность жизни

1. Однолетник
2. Двулетник
3. Многолетник с укороченным периодом жизни

- | | |
|--------------------------------|------------------------------------|
| 4. Многолетник | Е. Способ питания |
| Г. Периодичность пло доношения | 1. Аутотроф |
| 1. Монокарпик | 2. Симбионт |
| 2. Поликарпик | 3. Сапрофит |
| 3. Ремонтантное | 4. Полупаразит |
| Д. Отношение к увлажнению | 5. Паразит |
| 1. Ксерофит | Ж. Продолжительность жизни листьев |
| 2. Мезофит | 1. Листопадное |
| 3. Психрофит | 2. Полувечнозеленое |
| 4. Гигрофит | 3. Вечнозеленое |
| 5. Гидрофит | |

Каждая из перечисленных семи шкал содержит от 3 до 8 баллов и отражает лишь один из семи признаков, которыми в совокупности характеризуется в данном случае жизненная форма растения.

Перечень шкал признаков и число баллов в любой из них могут быть при необходимости сокращены или расширены. Например, в некоторых случаях полезно учесть признак Раункиера, или степень защищенности почек возобновления, для чего его можно закодировать следующей шкалой:

3. Признак Раункиера

- | | |
|--------------|------------------|
| 1. Фанерофит | 3. Гемикриптофит |
| 2. Хамефит | 4. Криптофит |

По таблицам (шкалам) А—Ж жизненную форму, например, остролистного клена в пределах его зоны комфорта можно выразить семизначным числом 1142211, что означает: дерево (1), прямостоячее (1), многолетник (4), поликарпик (2), мезофит (2), аутотроф (1), листопадное (1).

В данном семизначном шифре жизненной формы каждый признак занимает один разряд и свое определенное место (А Б В Г Д Е Ж). В восьмизначном шифре (А Б В Г Д Е Ж З) жизненная форма остролистного клена с учетом признака Раункиера передается числом из восьми разрядов 11422111, где последняя цифра (1) означает: фанерофит. При отсутствии сведений по любому признаку на его месте в составном символе ставится нуль (0).

Если некоторый признак должен быть закодирован больше, чем девятью баллами, то для него следует отвести два или больше разрядов в составном символе. При двух разрядах признак можно характеризовать в зависимости от степени его выраженности 99 градациями, или баллами, шкалы, при трех разрядах — 999 баллами.

Предлагаемая система жизненных форм и их обозначений удобна при любой и особенно перфокартной и машинной форме хранения и поиска информации, а также позволяет применить для обработки последней математический аппарат, в частности векторное исчисление и матричную алгебру. По этому принципу можно построить многопризнаковые шкалы и для любых других комплексов признаков, объединяемых каким-либо общим понятием, которое затем кодируется составным символом.

§ 5.03. Таксономический анализ Смирнова

Проблема классификации объектов не только в ботанике, но и во многих других областях современной науки и техники становится все более актуальной. Для решения этой задачи в настоящее время во многих странах затрачиваются значительные усилия. Судя по таким результатам, как, например, создание метода корреляционных плеяд (Терентьев, 1959) в биологии и кластер-анализа для классификации химических соединений, можно надеяться, что проблема биологической систематики также начнет приближаться к своему решению в обозримом будущем. Несомненно, что рабочей основой процесса систематизации будет специализированный, довольно сложный математический аппарат. Однако вполне могут быть полезны и актуальны как в настоящем, так и в будущем, наряду со сложными хотя и не столь совершенные, но зато и более простые методы таксономического анализа.

Теоретической предпосылкой одного из таких методов, разработанного энтомологом Е. С. Смирновым (1969), служит принцип неравноценности признаков в систематике, в отличие от принципа их равноценности, или адансоновского принципа, по имени французского ботаника Адансона (1727—1806). В настоящее время подавляющее большинство биологов-систематиков придерживается естественной системы при классификации биологических объектов, в которой принимается за основу неравнозначность признаков, иерархичность таксонов, отражающая эволюционно сложившиеся, филогенетические связи организмов. Степень таксономической близости видов или других систематических подразделений в естественной системе предполагается эквивалентной общности пройденного ими эволюционного пути. Чем реже встречается признак в совокупности объектов, тем большую ценность он представляет для их практической систематики. Именно это обстоятельство используется при построении таксономического отношения, которое в своей основе есть средний вес совпадающих признаков и тем самым отражает степень таксономической близости двух видов. Так как за вес признака принимается величина, обратная частоте его встречаемости, то понятно, что редкий признак получит больший вес, чем более распространенный.

Несмотря на подробное теоретическое обоснование и сравнительно простое, доступное изложение, из работы Е. С. Смирнова без дополнительного специального изучения затруднительно получить алгоритмы (т. е. расчетные схемы) таксономического анализа, по которым можно было бы сразу приступить к практическому его применению. Попытка построения таких алгоритмов с некоторой модификацией формул Е. С. Смирнова делается в настоящей работе. На следующем примере ознакомимся с расчетными схемами таксономического анализа.

Таксономическое отношение. В соответствии с удачным предложением Е. С. Смирнова введем понятие свойства, а признаками будем считать варианты этого свойства. Например, окраска цвет-

ков будет свойством, а ее варианты: синяя, фиолетовая, белая и другие будут признаками данного свойства.

Сравним два вида x и y из одного рода (в котором насчитывается 10 видов) по двум свойствам: по длине черешка листьев (D) и окраске венчика цветков (E). По свойству D листья у видов данного рода разделяются на три признака: листья сидячие (D_1), листья на черешках от 0 до 2 см (D_2) и листья на черешках длиной 2 см и больше (D_3). Окраска венчика цветка у видов рассматриваемого рода бывает синяя (E_1), фиолетовая (E_2), белая (E_3) и красная (E_4), т. е. свойство E разделяется на четыре признака. Всего по обоим свойствам $n = 7$ признаков; число видов $s = 10$. Таксономическое отношение подсчитывается в следующем порядке.

1. Распределим имеющиеся 10 видов по признакам обоих свойств. У шести видов листья сидячие ($6D_1$), у двух видов длина черешка от 0 до 2 см ($2D_2$), у остальных двух видов длина черешка больше 2 см ($2D_3$). Так же распределим эти виды по окраске цветка, получим два следующих распределения:

$$6D_1 + 2D_2 + 2D_3 = 10, \quad (a)$$

$$5E_1 + 3E_2 + 1E_3 + 1E_4 = 10. \quad (б)$$

Число видов при каждой букве называется фреквенцией соответствующего признака.

2. Выясним, по каким признакам сравниваемые два вида совпадают. Условимся, что наличие признака (или положительный признак) будет обозначаться большой буквой (например, D_1), а отсутствие признака (или отрицательный признак) соответствующей

Таблица 5.11. К вычислению таксономического отношения

Отношение признаков	Признаки видов		Фреквенция совпадающего признака, φ_i	Вес признака, $\omega_i = \frac{10 - \varphi_i}{\varphi_i}$	$\frac{1}{\varphi_i}$
		y			
Совпадающие положительные	D_1	D_1	$6D_1$	$4/6$	0,167
Совпадающие отрицательные	d_2	d_2	$8d_2$	$2/8$	0,125
	d_3	d_3	$8d_3$	$2/8$	0,125
	e_1	e_1	$5e_1$	$5/5$	0,200
	e_4	e_4	$9e_4$	$1/9$	0,111
Несовпадающие	E_2	e_2		—1	
	e_3	E_3		—1	
				0,278	0,728

малой буквой (например, d_1). Составим табл. 5.11, в которой отметим в соответствии с этим условием в столбцах 2, 3 совпадение и несовпадение признаков. Из таблицы видно, что у обоих видов листья сидячие (D_1), однако окраска венчика у вида x фиолетовая (E_2), а у вида y — белая (E_3), по остальным признакам, вернее по их отсутствию, виды также совпадают. Следовательно, виды x и y совпадают по одному положительному признаку (D_1) и по четырем отрицательным (d_2, d_3, e_1, e_4) и не совпадают по признакам $E_2 - e_2, e_3 - E_3$.

3. Для заполнения столбца 4 таблицы обратимся к распределениям (а) и (б), где записаны фреквенции положительных (имеющихся) признаков, которые непосредственно оттуда переносятся в столбец 4. Положительных совпадающих признаков всего один: D_1 , фреквенция которого $6D_1$ записывается в таблицу. Фреквенции совпадающих отрицательных признаков получаем вычитанием величины соответствующей положительной фреквенции из числа видов; так фреквенция для d_2 равна: $10 - 2 = 8$, т. е. фреквенцию при D_2 , которая равна 2, вычитаем из числа видов, равного 10. Иначе говоря, фреквенция отрицательного признака равна дополнению до общего числа видов для соответствующей фреквенции положительного признака. В разделе 3 табл. 5.11 в столбце 4 ничего не записываем, так как фреквенции несовпадающих признаков не участвуют в дальнейших расчетах.

4. Найдем вес положительных и отрицательных совпадающих, а также вес несовпадающих признаков. Вес положительного и отрицательного признака после указанного в пункте 3 приведения фреквенций равен:

$$\omega_i = \frac{s - \varphi_i}{\varphi_i}, \quad (5.06)$$

где ω_i — вес таксономического признака; s — общее число видов в анализируемом таксоне; φ_i — фреквенция положительного или отрицательного признака.

Вес пары несовпадающих признаков всегда равен единице с минусом. В нашем примере, учитывая фреквенции из столбца 4, найдем следующие веса по формуле (5.06): $\omega(D_1) = 4/6$, $\omega(d_2) = 2/8$, $\omega(d_3) = 2/8$; $\omega(e_1) = 5/5$, $\omega(e_4) = 1/9$, $\omega(E_2, e) = -1$, $\omega(e_3, E_3) = -1$, которые и записаны в столбце 5 табл. 5.11.

5. Вычислим таксономическое отношение для двух видов по формуле:

$$t_{xy} = \frac{1}{n} \sum_n \omega_i, \quad (5.07)$$

где t_{xy} — таксономическое отношение двух видов x и y ; n — число признаков, по которым они сравниваются; ω_i — веса совпадающих положительных и отрицательных, а также несовпадающих признаков;

$$t_{xy} = 1/7 (4/6 + 2/8 + 2/8 + 5/5 + 1/9 - 1 - 1) = 0,04.$$

6. Для таксономического анализа в пределах рода или другой группы, достаточно располагать относительными величинами таксономических отношений t_{xy} , получаемыми по формуле (5.07). Однако при сопоставлении результатов по двум или более таксонам, имеющим разное число видов, величины t_{xy} по этой формуле для данной цели непригодны, так как сильно зависят от числа видов. Поэтому лучше относительное таксономическое отношение сразу переводить в нормированное по формуле:

$$T_{xy} = \frac{2(t_{xy} + 1)}{s}, \quad (5.08)$$

где T_{xy} — нормированное таксономическое отношение; t_{xy} — относительное таксономическое отношение; s — число видов в анализируемом таксоне.

В рассматриваемом примере по формуле (5.08):

$$T_{xy} = \frac{2(0,04 + 1)}{10} = 0,208.$$

Величины нормированного таксономического отношения по формуле (5.08) от числа видов в анализируемом таксоне не зависят и поэтому предпочтительнее для применения. Величина T_{xy} всегда находится в пределах $1 > T_{xy} \geq 0$. Чем больше величина t_{xy} (или T_{xy}), тем больше степень таксономической близости между двумя видами.

7. Вычисление таксономического отношения по формуле (5.07), в которой участвуют веса и несовпадающих признаков, здесь приведено для иллюстрации принципа построения этого показателя. В практических расчетах более целесообразно пользоваться формулой

$$t_{xy} = \left(\frac{s}{n} \sum_f \frac{1}{\varphi_i} \right) - 1, \quad (5.09)$$

где t_{xy} — таксономическое отношение; s — общее число видов в данном таксоне; f — число совпадающих признаков; φ_i — фреквенция совпадающего признака.

В формуле (5.09) в отличие от формулы (5.07) фреквенции несовпадающих признаков не участвуют, в ней также не требуется вычислять веса признаков по формуле (5.06), а сразу вычисляются обратные величины фреквенций: $1/6 = 0,167$; $1/8 = 0,125$ и т. д., которые записаны в столбце 6, их сумма равна 0,728.

В нашем примере $s = 10$, $n = 7$, $f = 5$ и по формуле (5.09):

$$t_{xy} = (10/7 \cdot 0,728) - 1 = 0,04,$$

т. е. результат совпадает с полученным по формуле (5.07).

Далее таксономическое отношение, вычисленное по формуле (5.09) при необходимости сравнений можно перевести в нормированное по формуле (5.08).

Если по условиям анализа полученные величины таксономических отношений далее потребуется сравнить с другими, вычислен-

ными для таксонов с иными объемами видов, то целесообразно сразу вычислять нормированные таксономические отношения по формуле:

$$T_{xy} = \frac{2 \sum_i \frac{1}{\varphi_i}}{n}, \quad (5.10)$$

где T_{xy} — нормированное (сравнимое с другими) таксономическое отношение; f — число совпадающих признаков; φ_i — фреквенции совпадающих признаков; n — число признаков.

Вычислим по данным рассматриваемого примера, сумма обратных величин фреквенций для которого приведена в табл. 5.11:

$$T_{xy} = \frac{2 \cdot 0,728}{7} = 0,208.$$

Величина ненормированного таксономического отношения изменяется в пределах: $(s - 2)/2 > t_{xy} \geq -1$.

Величина нормированного таксономического отношения изменяется в пределах: $1 > T_{xy} \geq 0$.

Величина фреквенции совпадающих признаков ограничена пределами: $s - 1 \geq \varphi_i \geq 1$.

Таксономическая оригинальность вида. В наборах признаков, характеризующих виды, участвуют как банальные, часто встречающиеся, так и более редкие для данного рода признаки.

Наличие у видов редких признаков существенно облегчает их классификацию и опознавание, напротив, виды, обладающие лишь банальными признаками, различаются несколько труднее. Е. С. Смирнов предложил количественный показатель, который позволяет объективно различать виды по степени их таксономической оригинальности, т. е. по участию редких признаков в их характеристиках.

Для вычисления степени оригинальности вида требуется составить распределение всех видов анализируемого таксона по образцу распределений (а), (б).

Степень таксономической оригинальности вычисляется по формуле:

$$T_{xx} = \frac{\left(\frac{s-1}{n} \sum_i \frac{1}{\varphi_i} \right) - 1}{s-2}, \quad (5.11)$$

где T_{xx} — нормированная (т. е. употребляемая в сравнениях с видами из других таксонов) степень оригинальности вида; s — число видов в анализируемом таксоне; n — число признаков; φ_i — фреквенция отдельного признака.

Выясним степень оригинальности вида x , признаки которого перечислены в столбце 2 табл. 5.11, откуда они перенесены в столбец 1 табл. 5.12.

Таблица 5.12. К вычислению степени таксономической оригинальности вида x

Признак	Фреквенция, φ_i	$\frac{1}{\varphi_i}$
D_1	6	0,167
d_2	8	0,125
d_3	8	0,125
e_1	5	0,200
e_4	9	0,111
E_2	3	0,333
e_3	9	0,111
		1,172

1. Фреквенции положительных признаков D_1 и E_2 без изменений переписываем в столбец 2 табл. 5.12 из распределений (а) и (б). Фреквенции отрицательных признаков: d_2, d_3, e_1, e_4, e_3 , так же как и при вычислении таксономического отношения, равны остаткам от вычитания из общего числа видов соответствующих фреквенций положительных признаков; так для d_2 фреквенция равна $10 - 2 = 8$, для d_3 тоже $= 8$, для e_1 она равна $10 - 5 = 5$.

2. В столбце 3 табл. 5.12 приводятся обратные величины фреквенций, сумма которых равна 1,172.

$$3. \text{ По формуле (5.11): } T_{xx} = \frac{\frac{10-1}{7} \cdot 1,172 - 1}{10-2} = 0,063.$$

Для сравнения вычислим по этой формуле степень оригинальности второго вида (y), признаки которого перечислены в столбце 3 табл. 5.11:

$$\Sigma 1/\varphi_i = 1/6 + 1/8 + 1/8 + 1/5 + 1/9 + 1/7 + 1/1 = 1,871;$$

$$T_{yy} = \frac{\frac{10-1}{7} \cdot 1,871 - 1}{10-2} = 0,176.$$

Степень оригинальности вида y больше, чем вида x . Это объясняется тем, что в наборе признаков вида x участвуют в основном признаки обычные, а в числе признаков вида y фигурирует редкий признак E_3 (его фреквенция равна 1), место которого в составе признаков вида x занимает банальный признак e_3 , с фреквенцией $10 - 1 = 9$, т. е. присутствующий у 9 видов из 10. Таким образом, чем больше величина T_{xx} , тем выше степень оригинальности вида.

Предельные значения рассматриваемого показателя определяются выражением:

$$1 > T_{xx} \geq 0,$$

т. е. это величина всегда положительная и не достигающая единицы, но асимптотически приближающаяся к ней.

Введенные Е. С. Смирновым коэффициенты T_{xx} и T_{xy} позволяют сравнивать таксономические отношения t_{xx}, t_{xy} некоторого таксона с таксономическими отношениями t_{xx}, t_{xy} других таксонов. Однако коэффициенты T_{xx}, T_{xy} не отражают связи между значениями t_{xx}, t_{xy} для одного и того же таксона, так как введением преобразования с использованием разных пределов для T_{xx} и T_{xy} получаем матрицу таксономических отношений, неэквивалентную матрице, полученной до преобразования. Чтобы сделать t_{xx} и t_{xy}

сравнимыми для различных таксонов и в то же время отразить связь между t_{xx} и t_{xy} для одного и того же таксона можно поступить следующим образом. Принимая во внимание неравенства ¹:

$$s - 1 \geq t_{xx} \geq \frac{1}{s - 1}, \quad (\text{в})$$

$$\frac{s - 2}{2} > t_{xy} \geq -1, \quad (\text{г})$$

делаем вывод, что таксономическое отношение, как внутривидовое, так и межвидовое, изменяется в пределах: $s - 1 \geq t \geq -1$, откуда: $t_{\min} = -1$; $t_{\max} - t_{\min} = s$.

Подставляя две последние величины в формулу:

$$T = \frac{t - t_{\min}}{t_{\max} - t_{\min}}, \quad (5.12)$$

получаем новый нормированный универсальный коэффициент: $\tau = (t + 1)/s$, который варьирует от 0 до 1, унифицирует выражение таксономических отношений t_{xx} , t_{xy} и сохраняет связь между t_{xx} и t_{xy} для одного и того же таксона.

* * *

Таксономический анализ можно применять не только к морфологическим признакам, но также и для анализа встречаемости видов какого-либо таксона на определенных экологических разностях, и в любом другом случае, когда виды некоторой полностью представленной систематической группировки возможно распределить по имеющимся признакам. Представляется перспективным применение таксономического анализа также во внутривидовой систематике, особенно при построении систем классификации сортов культивируемых растений, где филогенетические взаимосвязи имеют меньшее значение, чем при систематике более крупных таксонов.

§ 5.04. Сравнение флор по видовому составу

При некоторых флористических и биогеографических исследованиях встречается необходимость сравнить один регион с другим, или построить матрицу сравнений видового состава различных флор или фаун. Для этого применяют так называемые коэффициенты сходства, из которых наиболее употребительны коэффициенты Жаккара:

$$K_I = \frac{c}{a + b - c}, \quad (5.13)$$

и коэффициент Чекановского:

$$K_C = \frac{2c}{a + b}, \quad (5.14)$$

¹ При дальнейшем изложении таксономического анализа мы пользовались консультацией П. В. Тамарина, которому выражаем свою благодарность.

где a — число видов в регионе A ; b — число видов в регионе B ; c — число видов, встречающихся одновременно в обоих регионах.

Однако у этих коэффициентов, как и у большинства других, служащих для цели сравнения флор, отсутствует математическое обоснование, для них не разработана методика вычисления ошибки и критерия достоверности различия, что сильно ограничивает возможности их применения. Коэффициенты (5.13) и (5.14) различаются по логике их построения. У коэффициента Жаккара множество совмещено у обоих регионов, а в коэффициенте Чекановского в качестве расчетных данных применяются два отдельных множества c , реально существующих каждое в своем регионе.

Из двух названных более обоснован принцип построения у коэффициента Чекановского, который представляет собой, как показано ниже, среднюю долю общих видов в двух регионах. В математической статистике при сравнении долей предварительно их усредняют, используя объемы выборок как веса по формулам (4.09), (4.10):

$$p' = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2},$$

$$p_1 = n_1/N_1, \quad p_2 = n_2/N_2,$$

где p' — средняя доля; N_1, N_2 — объемы первой и второй выборок; p_1, p_2 — доли наблюдаемого признака в первой и второй выборках; n_1, n_2 — фактические численности вариант с наблюдаемым признаком.

Применяя те же обозначения, что и к формулам (5.13) и (5.14), взвешенную среднюю долю общих видов в регионах A и B вычислим следующим образом:

$$p_1 = c/a, \quad p_2 = c/b, \quad p' = \frac{\frac{ac}{a} + \frac{bc}{b}}{a + b} = \frac{2c}{a + b},$$

что и представляет собой коэффициент Чекановского.

При распределении дат, близкому к нормальному, сравнивать видовой состав флор (фаун) рекомендуется по формуле:

$$F = \frac{(p_1 - p_2)^2}{\left(\frac{1}{a} + \frac{1}{b}\right)(1 - K_C) K_C}, \quad (5.15)$$

где F — критерий Фишера; p_1, p_2 — доли общих видов в двух сравниваемых регионах; a, b — число видов в первом и втором регионах; K_C — средняя доля, или коэффициент Чекановского по формуле (5.14). Число степеней свободы для оценки достоверности критерия Фишера, определяется из выражений: $\nu(1) = 1$; $\nu(2) = a + b - 2$.

Сравним флористический район A , в котором насчитали $a = 1000$ видов, с районом B , в котором было $b = 500$ видов, общих для обоих районов оказалось: $c = 100$ видов.

1. Доля общих видов в районе A :

$$p_1 = c/a = 100/1000 = 0,1.$$

2. Доля общих видов в районе B :

$$p_2 = c/b = 100/500 = 0,2.$$

3. Коэффициент сходства, или средняя доля, по формуле (5.14):

$$K_c = \frac{2 \cdot 100}{1000 + 500} = 0,133.$$

4. Оцениваем достоверность различия по составу флоры двух сравниваемых районов по формуле (5.15):

$$F = \frac{(0,1 - 0,2)^2}{\left(\frac{1}{1000} + \frac{1}{500}\right)(1 - 0,133)0,133} = 28,8.$$

При $\nu(1) = 1$ и $\nu(2) = 1000 + 500 - 2 = 1498$ по табл. 9П критических значений критерия Фишера $F'(0,95) = 3,84$.

Вычисленное значение критерия значительно превышает табличное: $28,8 > 3,84$, поэтому два сравниваемых района A и B следует считать достоверно различающимися по доле общих для них видов, т. е. они не схожи по флористическому составу. При изучении флористических взаимосвязей можно применять также показатели связи, употребляемые в математической статистике для четырехпольных таблиц распределения частот.

Попытаемся выразить степень флористической близости регионов при помощи коэффициента сходства, построенного на основе коэффициента взаимной сопряженности Чупрова:

$$K_4 = \frac{|c(a-b)| - \frac{a+b}{2}}{\sqrt{2abc(a+b-2c)}}, \quad (5.16)$$

или в логарифмированном виде:

$$K_4 = \exp \left\{ \ln \left| |c(a-b)| - \frac{a+b}{2} \right| - \frac{\ln 2 + \ln a + \ln b + \ln c + \ln(a+b-2c)}{2} \right\},$$

с оценкой по критерию хи-квадрат:

$$\chi^2 = K_4^2(a+b),$$

где K_4 — коэффициент флористического сходства; a, b — численность видов в регионах A и B ; c — число общих для них видов; χ^2 — критерий хи-квадрат, величина которого оценивается при $\nu = 1$, т. е. при одной степени свободы.

Необходимость в логарифмировании формулы (5.16) возникает в связи с тем, что нередко при флористических анализах число сравниваемых видов бывает довольно велико, и при их умножении подкоренное выражение становится слишком большим, что не удобно при вычислении.

В четырехклеточной табл. 5.13 показан один из вариантов размещения частот при оценке флористического сходства. В этой таб-

Таблица 5.13. Размещение частот при оценке флористического сходства

Флористиче- ские район	Состав флор		Σ
	число раз- личающихся видов	число сов- падающих видов	
A	$a - c$	c	a
B	$b - c$	c	b
Σ	$a + b - 2c$	$2c$	$a + b$

лице можно менять местами строки и столбцы, а также написать строки на место столбцов и, наоборот, результаты от этого не изменяются. Направление связи в соответствующих конкретных случаях можно определить по неравенству: $a > b$, когда коэффициент K_4 будет положительным, или $a < b$, когда он будет отрицательным. Однако в общем случае знак коэффициента не имеет принципиального значения, поскольку можно менять местами обозначения

регионов. Коэффициент сходства и коэффициент корреляции для четырехклеточных таблиц (и только для них) совпадают по величине.

Найдем по формуле (5.16) через логарифмы коэффициент сходства для вышеприведенных данных: $a = 1000$, $b = 500$, $c = 100$ видов.

$$K_4 = \exp \left\{ \ln \left| \left| 100 (1000 - 500) \right| - \frac{1000 + 500}{2} \right| - \frac{\ln 2 + \ln 1000 + \ln 500 + \ln 100 + \ln (1000 + 500 - 2 \cdot 100)}{2} \right\} = 0,136.$$

$\chi^2 = 0,136^2 (1000 + 500) = 27,9$, что больше критического значения при любой стандартной доверительной вероятности. Следовательно, коэффициент корреляции, или сходства, достоверен. Однако степень сходства, отражаемая величиной коэффициента K_4 , в данном случае мала. Это свидетельствует о том, что два флористических региона A и B скорее различаются, чем сходны по доле общих видов.

Проверим условия нормальности, которые для оценки неизвестной вероятности следующие: $N'p' > 10$; $N'(1 - p') > 10$; $N' > 200$ или $N'p'(1 - p') > 9$, где N' — общий объем совокупности; p' — усредненная доля. Для рассматриваемых данных: $N' = 1500$; $p' = 0,133$; $1500 \cdot 0,133 = 199 > 10$; $1500 \cdot 0,867 = 1300 > 10$; $1500 > 200$; $1500 \cdot 0,133 \cdot 0,867 = 173 > 9$. Следовательно, требования условий нормальности выборки в данном случае удовлетворяются и выводы верны.

Кроме того, сравнение тех же долей по формуле (4.11), свободной от соблюдения условия нормальности выборок, также дает результат, аналогичный полученному выше по формуле (5.15):

$$F = (\varphi_2 - \varphi_1)^2 \frac{N_1 N_2}{N_1 + N_2} = (0,927 - 0,644)^2 \frac{1000 \cdot 500}{1000 + 500} = 26,6,$$

стандартное значение: $F_\infty(0,95) = 3,84$, где φ_2 и φ_1 соответствуют преобразованным через арксинус долям $p_2 = 0,2$ и $p_1 = 0,1$.

Возможно, однако, что коэффициент корреляции и оценка его по хи-квадрат при четырехпольном распределении ограничены в

своих возможностях малым числом степеней свободы. Поэтому с равным те же доли по формуле (4.17) для четырехпольного распределения факториалов частот в применении к обозначениям в табл. 5.13:

$$P(\chi^2) = \frac{a!b!(a+b-2c)!(2c)!}{(a-c)!c!(b-c)!c!(a+b)!} = \frac{1000!500!1300!200!}{900!100!400!100!1500!};$$

$$\lg P(\chi^2) = -23,4796869.$$

Логарифмы факториалов чисел 1300 и 1500 здесь вычислены при помощи формулы Стирлинга:

$$\lg n! = (n + 0,5) \lg n - n \lg e + \lg \sqrt{2\pi},$$

где n — число, факториал которого вычисляется.

Как видим, вероятность получить одинаковые частоты в клетках табл. 5.13 почти равна нулю. Следовательно, два сравниваемых региона A и B существенно различаются по доле общих видов.

Множественные взаимосвязи флор можно выявить и при помощи полихорического коэффициента сопряженности (§ 2.12). При этом рекомендуется: а) применять известные, достаточно математически обоснованные методы и критерии математической статистики вместо заменяющих их по сути дела суррогатов в виде различных коэффициентов сходства; б) ориентироваться и на понятие различия регионов, а не только на их сходство.

Общим недостатком существующих коэффициентов флористической связи считают также то, что они учитывают более качественную сторону сходства, чем количественную, так как на их величину не оказывает влияние площадь распространения видов в сравниваемых регионах. Однако с таким же правом от коэффициентов флористического сходства можно потребовать, чтобы они отражали дизъюнктивный или конъюнктивный, прогрессивный или регрессивный характер границ ареалов видов, что имеет большое значение при флористическом анализе. Важно также учитывать рельеф сравниваемых регионов, качественное состояние растений, например бонитет у древесных видов. Увеличение числа сравниваемых признаков приведет нас и в данном случае к еще недостаточно разрешенным общебиологическим проблемам многопризнаковой классификации и структуры многофакторных комплексов. Поэтому до разрешения указанных проблем пока нецелесообразно вводить в процедуру флористического сравнения дополнительную количественную характеристику и увеличивать число учитываемых при сравнении факторов.

§ 5.05. Логистический анализ

При помощи этого метода возможно исследовать лишь те экспериментальные зависимости, форма которых графически близка к логистической кривой в ее двух модификациях (§ 3.11). Несмотря на такое ограничение, круг вопросов, которые могут быть изучены с помощью логистического анализа, остается очень широким. До-

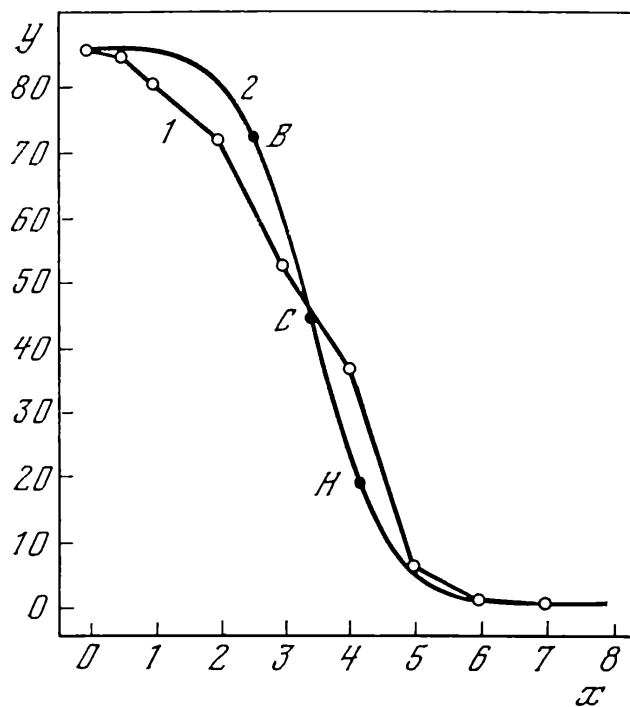


Рис. 31. Эмпирическая (1) и логистическая (2) кривые зависимости всхожести семян жимолости (y) от числа лет их хранения (x)

популяции, при их отмирании и в различных переходных состояниях. Логистической закономерности присуще свойство отражать изменение возрастающего ускорения процесса на замедляющееся, или наоборот — при обратной форме кривой. Эта важная особенность дает возможность находить аналитическим путем различные критические или оптимальные и другие практически ценные точки, в том числе и так называемую эффективную дозу (LD_{50} , или E_{50}). Следовательно, логистический анализ дает возможность аналитически интерпретировать и графически аппроксимировать некоторый процесс и, кроме того, включает в себя то, что получают обычно при помощи пробит-анализа, т. е. определение эффективной дозы.

В данном параграфе излагается наиболее простой по технике вычислений метод логистического анализа с визуальной аппроксимацией по прямой линии. От прочих методов вычисления коэффициентов логистического уравнения, приведенных в литературе, данный метод отличается тем, что коэффициенты уравнения определяются при помощи двух избранных точек, произвольно взятых на прямой. Прямая линия проводится от руки и представляет собой зависимость натуральных логарифмов несколько преобразованных значений функции от исходных величин аргумента. Натуральные логарифмы в данном случае дают возможность обойтись без двукратного преобразования их естественной формы в искусственную и обратно, как это пришлось бы делать, используя логарифмы с основанием 10. Применение способа избранных точек при отыскании коэффициентов уравнения прямой линии позволяет значительно

статочко сказать, что логистическая кривая — это закон роста — явления присущего всем известным формам и уровням жизни. По логистическому закону происходит увеличение как массы особи и ее отдельных органов, так и численности различных популяций. Обратный процесс: уменьшение численности популяций по эндогенным или экзогенным причинам, понижение жизнеспособности особей и их органов, также происходит по обратной логистической кривой (рис. 31).

В целом логистический закон отражает динамику многих жизненных процессов в пространстве и времени как в прогрессивном, так и в регрессивном направлениях, при зарождении нового организма или

Таблица 5.14. К вычислению коэффициентов логистического уравнения по методу визуальной аппроксимации

Срок хранения семян, лет, x	Всхожесть, %, y	$\frac{a_1}{y}$	$\frac{a_1}{y} - 1$	
1	2	3	4	5
7	0,1	863	862	6,759
6	0,64	135	134	4,898
5	6	14,4	13,4	2,595
4	36	2,4	1,4	0,3365
3	52	1,66	0,66	—0,4155
2	71,4	1,21	0,21	—1,5606
1	80,3	1,075	0,075	—2,5903
0,5	85	1,015	0,015	—4,2
0,1	86	1,003	0,003	—5,809

снизить трудоемкость расчетов без существенного снижения их эффективности, так как несмотря на указанные упрощения в алгоритме расчетов, точность излагаемого метода вполне достаточна для большинства биологических исследований при условии, что исходные данные достаточно близки к логистической зависимости.

Техника расчетов видна из рассматриваемого примера аппроксимации логистической кривой зависимости всхожести семян 23 видов жимолости от срока их хранения. Исходные данные приведены в первых двух столбцах табл. 5.14. Проведем аппроксимацию этой зависимости логистической кривой по уравнению (3.106):

$$y' = \frac{a_1}{1 + e^{\gamma + \beta x}} + a_0,$$

где y' — теоретические значения функции; a_1, a_0 — верхняя и нижняя асимптоты кривой (см. § 3.11); $e = 2,7183$ — основание натуральных логарифмов; γ, β — коэффициенты логистического уравнения; x — аргумент.

Попытаемся найти также ответы на следующие практические вопросы: какова критическая продолжительность хранения семян жимолости, далее которой их высевать нецелесообразно, какова при этом критическом сроке всхожесть семян и, наконец, какова в среднем максимальная теоретическая всхожесть семян жимолости.

1. В соответствии с графиком эмпирической линии регрессии выбираем значение верхней асимптоты $a_1 = 86,3\%$ (см. рис. 31), а также значение нижней асимптоты $a_0 = 0\%$. В данном конкретном случае эти две величины обозначают верхний максимальный предел всхожести семян жимолости и нижний минимальный предел, который, естественно, равен 0% .

2. Значение верхней асимптоты делим поочередно на все значения y (столбец 3) и вычитаем из них единицу (столбец 4).

3. Логарифмируем числа столбца 4 (столбец 5).

Для этого пользуемся натуральными логарифмами, которые можно найти или в таблицах e^x, e^{-x} , или в специальных таблицах на-

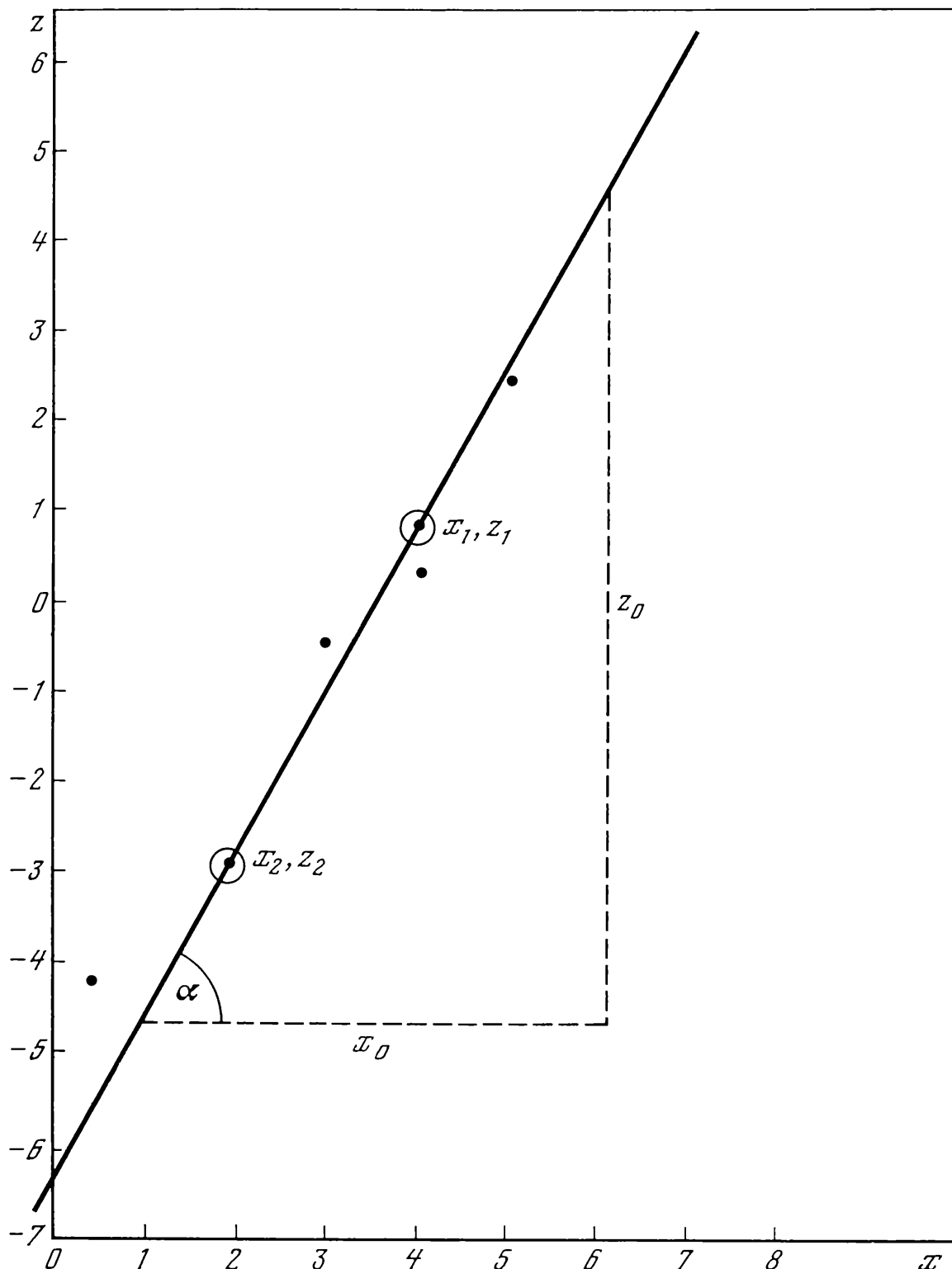


Рис. 32. Выравнивание прямой линией логистической зависимости всхожести семян (z — преобразованные величины функции) от x — числа лет их хранения x_0, z_0 — построение отрезков для определения угла α ; x_1z_1, x_2z_2 — произвольные точки на прямой, α — угол наклона прямой линии. Эмпирические данные на графике показаны точками

туральных логарифмов, или (что менее удобно) путем преобразования десятичных логарифмов в натуральные по формуле

$$\ln x = 2,30259 \lg x \text{ или } \ln x = \lg x / 0,43429.$$

4. Числа из столбца 5 обозначим z и нанесем их на ось ординат рис. 32, по оси абсцисс которого отложены значения x — число

лет хранения семян до момента их проращивания. Полученные точки при помощи линейки выравниваем прямой линией, на которой произвольно отмечаем две точки с координатами x_1z_1 и x_2z_2 .

5. Находим на осях соответствующие численные значения координат: $x_1 = 4$; $z_1 = 1$; $x_2 = 2$, $z_2 = -2,7$. Обратим внимание на то, что в отрицательной части значений z на оси ординат они по абсолютной величине возрастают сверху вниз, в том числе и в промежутках между отмеченными значениями.

6. Вычисляем величины коэффициентов логистического уравнения по формулам:

$$\gamma = \frac{x_1z_2 - x_2z_1}{x_1 - x_2}, \quad \beta = \frac{z_2 - z_1}{x_2 - x_1}, \quad (5.17)$$

$$\gamma = [4(-2,7) - 2 \cdot 1]/(4 - 2) = -6,4; \quad \beta = (-2,7 - 1)/(2 - 4) = 1,85.$$

Проверку вычислений проводим по формулам:

$$\gamma = z_1 - \beta x_1; \quad \beta = (z_1 - \gamma)/x_1,$$

$$\gamma = 1 - 1,85 \cdot 4 = -6,4; \quad \beta = [1 - (-6,4)]/4 = 1,85. \quad (5.18)$$

Формулы (5.17) и (5.18) получены из системы уравнений прямой линии, составленных для двух избранных точек: $\gamma + \beta x_1 = z_1$; $\gamma + \beta x_2 = z_2$, из которых и непосредственно совместным их решением также можно получить искомые коэффициенты.

Укажем также и третий наиболее простой способ решения: величина коэффициента $\gamma = -6,4$ равна величине отрезка на оси ординат, отсекаемого прямой линией, а коэффициент β равен тангенсу угла α , т. е. отношению длины отрезков $z_0/x_0 = 1,85$. Отрезки z_0 , x_0 следует измерять в единицах длины по их фактической протяженности на графике. Измерив угол α транспортиром, можем найти также его тангенс по таблицам: $\text{tg } 61^\circ 30' = +1,85 = \beta$. Знак тангенса в таблицах при этом в пределах 180° также соответствует знаку коэффициента β , если угол отсчитывать от оси абсцисс по направлению против движения часовой стрелки. Указанные методы нахождения обоих коэффициентов можно применять и одновременно, для взаимного контроля результатов.

Для вычисления теоретических значений функции, т. е. всхожести семян после определенного срока их хранения по уравнению логистической зависимости:

$$y' = \frac{86,3}{1 + e^{-6,4 + 1,85x}},$$

составим табл. 5.15, в которой проведены следующие вычисления.

1. Значения аргумента возьмем в несколько большем диапазоне: от 0 до 8 лет с шагом через 0,5 года для того, чтобы получить соответствующее число значений функции и обеспечить достаточным числом точек теоретическую кривую (столбец 1).

2. В соответствии с уравнением умножаем числа столбца 1 на коэффициент $\beta = 1,85$ и вычитаем коэффициент $\gamma = -6,4$ (столбцы 2, 3).

Таблица 5.15. Вычисление теоретических значений логистической функции

	βx	$\gamma + \beta x$	$e^{\gamma + \beta x}$	$1 + e^{\gamma + \beta x}$	y'	y	$(y' - y)^2$
1	2	3	4	5	6	7	8
8	14,8	8,4	4447	4448	0,0194		
7,5	13,87	7,47	1755	1756	0,049		
7	12,95	6,55	699,24	700,24	0,123	0,1	0
6,5	12,03	5,63	278,7	279,7	0,309		
6	11,1	4,7	109,9	110,9	0,778	0,64	0,0256
5,5	10,17	3,77	43,38	44,38	1,94		
5	9,25	2,85	17,29	18,29	4,72	6	1,69
4,5	8,32	1,92	6,82	7,82	11,04		
4	7,4	1,0	2,72	3,72	23,2	36	163,84
3,5	6,47	0,07	1,073	2,073	41,6		
3	5,55	-0,85	0,43	1,43	60,3	52	68,89
2,5	4,62	-1,78	0,17	1,17	73,8		
2	3,7	-2,7	0,067	1,067	80,9	71,4	90,25
1,5	2,77	-3,63	0,026	1,026	84,1		
1	1,85	-4,55	0,011	1,011	85,4	80,3	26,01
0,5	0,925	-5,475	0,004	1,004	86,0	85,0	1,0
0	0	-6,4	0,002	1,002	86,1		
							351,7

3. Получившиеся в столбце 3 значения есть натуральные логарифмы, по которым следует найти соответствующие им числа, для чего воспользуемся или таблицами e^x , e^{-x} , или таблицами натуральных логарифмов. В результате потенцирования получим числа столбца 4, прибавив к ним единицу, найдем полное значение знаменателя логистического уравнения (столбец 5).

4. Разделив значение верхней асимптоты $a_1 = 86,3$ поочередно на числа столбца 5, получим теоретические значения функции y' (столбец 6), по которым построим соответствующую логистическую кривую (см. рис. 31,2).

5. Для вычисления ошибки уравнения сумму квадратов разностей эмпирических и теоретических значений функции: $\sum (y' - y)^2 = 351,7$ подставляем в формулу (3.04):

$$m_{y \cdot x} = \sqrt{\frac{351,7}{7-3}} = 9,4\%.$$

Как видно из табл. 5.15, число разностей эмпирических и теоретических значений функции $N = 7$, а число параметров логистического уравнения $n = 3$, что и отражено в только что приведенном расчете ошибки прогноза.

6. Анализ полученного уравнения логистической кривой показывает, что точка ее пересечения с осью ординат по формуле (3.111) равна:

$$y_{\Pi} = \frac{86,3}{1 + e^{-6,4}} = 86,1\%,$$

что соответствует в среднем для рода жимолость верхнему пределу всхожести семян. Координаты точки с перегиба кривой по формулам (3.112) равны:

$$x_c = -(-6,4/1,85) = 3,5 \text{ года}, \quad y_c = 86,3/2 = 43\%,$$

что соответствует критическому сроку хранения семян жимолости, после чего их всхожесть начнет быстро падать. Критическое значение всхожести при этом равно 43%, ниже которой она становится все более недостаточной для посева.

Более детальный анализ уравнения логистической кривой проводится определением координат следующих дополнительных критических точек (см. рис. 31):

$$x_B = \frac{-\gamma + \ln(2 - \sqrt{3})}{\beta}, \quad (5.19)$$

$$x_H = \frac{-\gamma + \ln(2 + \sqrt{3})}{\beta}, \quad (5.20)$$

$$y_B = \frac{a_1}{3 - \sqrt{3}} + a_0, \quad (5.21)$$

$$y_H = \frac{a_1}{3 + \sqrt{3}} + a_0, \quad (5.22)$$

где $\ln(2 - \sqrt{3}) = -1,317$; $\ln(2 + \sqrt{3}) = 1,317$; $3 - \sqrt{3} = 1,268$; $3 + \sqrt{3} = 4,732$; $x_B = (6,4 - 1,317)/1,85 = 2,75$; $x_H = (6,4 + 1,317)/1,85 = 4,17$; $y_B = 86,3/1,268 + 0 = 68,06$; $y_H = 86,3/4,732 + 0 = 18,24$.

Полученные величины означают, что потеря всхожести семян особенно интенсивно происходит, начиная с $x_B = 2,7$ года до $x_H = 4,2$ года, в это время всхожесть падает с $y_B = 68,1\%$ до $y_H = 18,2\%$, после чего падение всхожести снова замедляется. Величины констант $1/(3 - \sqrt{3})$, $1/2$, $1/(3 + \sqrt{3})$ в последних формулах, равные соответственно 0,79; 0,5 и 0,21, означают, что в точках В, С, Н кривой всегда, независимо от прочих условий, будет сохраняться всхожесть на 79, 50 и 21% от начальной, что происходит в данном случае через 2,7; 3,5 и 4,2 года после начала хранения.

* * *

Элементарный анализ логистической зависимости может быть проведен также и просто по графику, с которого часто нетрудно снять приблизительные численные значения верхней (a_1) и нижней (a_0) асимптот, затем выровнять эмпирические точки от руки плавной кривой. Определив по формуле $y_a = a_1/2 + a_0$ ординату точки перегиба (y_a), опустим из последней перпендикуляр на горизонтальную ось и найдем абсциссу точки перегиба (x_a), что даст приблизительное значение эффективной дозы, или критической точки данного процесса или явления.

В некоторых случаях требуется получить также и значение точки пересечения логистической кривой с осью абсцисс, так в

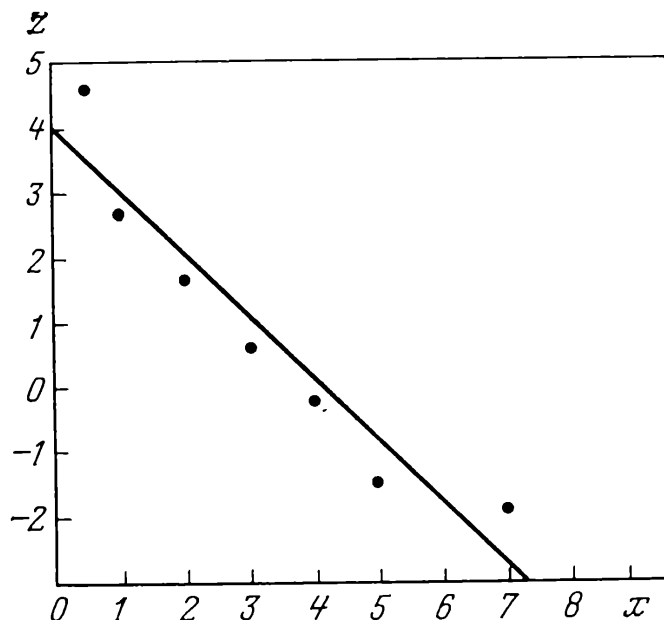


Рис. 33. Выравнивание прямой линией процента невсхожих семян жимолости (z — преобразованные величины функции) в зависимости от числа лет хранения (x)

Эмпирические данные на графике показаны точками

Таблица 5.16. К определению точки пересечения линии логистической функции с осью абсцисс

Срок хранения семян, лет, x	Доля невсхожих семян, %, y	$\frac{a_1}{y - a_0} - 1$	
7	99,9	0,16	-1,83
6	99,36	0,17	-1,77
5	94,0	0,25	-1,39
4	64,0	1,00	0
3	48,0	1,94	0,66
2	28,6	5,85	1,77
1	19,7	16,5	2,8
0,5	15,0	99,0	4,6
0,1	14,0	—	—

$$a_0 = 14\% \quad a_1 = 100\%$$

нашем примере это помогло бы приблизительно оценить максимальный срок сохранения семенами жимолости хотя бы и очень небольшой всхожести. В большинстве случаев эту точку пересечения трудно найти даже приблизительно ввиду асимптотического приближения кривой к горизонтальной оси. Тем не менее найти точку пересечения логистической кривой с осью абсцисс возможно, для чего потребуются следующие несложные преобразования исходных данных. Сначала получим дополнения к значениям функции. Так, в нашем примере следует взять в качестве функции процент невсхожих семян (вместо проросших), что составляет дополнение к 100% по отношению к первоначальным величинам функции (см. столбец 2 табл. 5.16), и биологически отражает динамику падения всхожести семян в зависимости от срока их хранения.

В столбце 3 таблицы из значений функции вычитается величина нижней асимптоты ($y - a_0$), затем на полученные числа делим величину верхней асимптоты и вычитаем из частного единицу. Для полученных чисел находим их натуральные логарифмы (z), которые откладываем на графике (рис. 33). Полученные точки выравниваем прямой линией, которая пересекает ось абсцисс в точке 7,3 года, что и можно считать теоретически максимальным сроком сохранения всхожести у семян жимолости, находящихся в помещении при комнатной температуре. Таким образом, максимальная всхожесть семян видов жимолости в среднем 86%, критический срок их хранения — 3,5 года, при этом всхожесть в среднем — 43%; полностью теряют всхожесть эти семена через 7 лет хранения. Приведенные числа можно считать практическими рекомендациями для семенного размножения видов жимолости.

Приведем также способ, который можно практически с достаточной точностью применить к прогнозу срока полной потери всхожести семян, что предоставит возможность значительно сократить время и труд при этом виде исследований. Для этого необходимо предварительно получить следующие исходные показатели.

1. Всхожесть семян в начале срока их хранения в процентах (y_1) и время (x_1) в месяцах или годах, прошедшее со времени сбора семян до момента первого определения всхожести. Так как непосредственно после сбора семян их обычно не проращивают, то максимальная всхожесть семян (a_0) должна быть хотя бы немного больше, чем при первом определении всхожести, т. е. $a_0 > y_1$.

2. Всхожесть (в процентах — y_2) после некоторого времени (x_2) их хранения. Продолжительность последнего периода зависит от скорости падения всхожести семян у конкретного вида. Следует стремиться к тому, чтобы x_2 было равно примерно половине того срока, после которого семена по предположению экспериментатора полностью потеряют всхожесть. Чем больше по продолжительности период x_2 и чем точнее найдена всхожесть y_2 , тем достовернее будет определен срок (x_3) полной потери всхожести (a_1). Для рассмотренного здесь фактического примера перечисленные величины равны: $a_0 = 86,3\%$; $y_1 = 86\%$; $x_1 = 0,1$ года; $x_2 = 3$ года; $y_2 = 52\%$; $a_1 = 0$. Расчеты коэффициентов логистического уравнения проводим по следующим формулам:

$$\beta = \frac{\ln\left(\frac{100 - a_1}{a_0 - y_2} - 1\right) - \ln\left(\frac{100 - a_1}{a_0 - y_1} - 1\right)}{x_2 - x_1}, \quad (5.23)$$

$$\gamma = \ln\left(\frac{100 - a_1}{a_0 - y_2} - 1\right) - \beta x_2,$$

подставляя в которые известные величины получим:

$$\beta = \frac{\ln\left(\frac{100 - 0}{86,3 - 52} - 1\right) - \ln\left(\frac{100 - 0}{86,3 - 86} - 1\right)}{3 - 0,1} = -1,778;$$

$$\gamma = \ln\left(\frac{100 - 0}{86,3 - 52} - 1\right) + 1,778 \cdot 3 = 5,984.$$

Координаты точки перегиба, или E_{50} , определим по формулам (3.112)

$$x_E = \left| \frac{\gamma}{\beta} \right| = \frac{5,9837}{1,778} = 3,37 \text{ года}, \quad y_E = \frac{a_0}{2} = \frac{86,3}{2} = 43,2\%.$$

Удваивая величину x_E , получим срок полной потери всхожести $x_3 = 2x_E = 2 \cdot 3,37 = 6,7$ года, что достаточно близко соответствует величине найденной выше (7 лет). Способ удвоения x_E для определения естественного конца какого-либо процесса можно применять лишь в том случае, если известны данные, которыми начинается процесс. В нашем случае известна начальная всхожесть семян $a_0 = 86,3\%$ и процесс падения всхожести описан с первоначальной стадии, поэтому можно применить правило симметрии начала и конца логистического процесса.

В заключение отметим, что все методы, рассмотренные здесь на примере анализа всхожести семян жимолости, можно применять к любым биологическим и небιологическим явлениям природы, процесс развития которых соответствует логистическому закону.

§ 5.06. Пробит-анализ

Методы пробит-анализа применяются исключительно лишь к таким зависимостям, эмпирические линии регрессии у которых похожи на логистические кривые. Поэтому, как уже указывалось (§ 5.05), результаты, вычисляемые обычно при помощи пробит-анализа, можно получить, наряду с другими данными, посредством логистического анализа, вследствие чего надобность в пробит-анализе вообще отпадает. Однако пробит-анализ имеет традиционную сферу применения в радиобиологии, токсикологии, иммунологии, фармакологии и некоторых других разделах наук, где проведение пробит-анализа связано с привычными методиками, имеющими важное научное и практическое значение. Результаты определения числового значения эффективной дозы (E_{50}) посредством логистического анализа и пробит-анализа различаются несущественно, поэтому тот и другой метод могут применяться в зависимости от желания исследователя. Вместе с тем отметим, что логистический анализ кроме основного результата — величины E_{50} дает возможность также построить более теоретически обоснованную линию регрессии (логистическую кривую) с вычислением коэффициентов ее уравнения, что создает необходимую предпосылку для дальнейшего математико-биологического анализа изучаемой зависимости и получения некоторых дополнительных данных.

Итак, основное и единственное назначение пробит-анализа заключается в определении смертельной (LD_{50}) или эффективной (E_{50}) дозы некоторого фактора, действующего на конкретные живые объекты. Методы пробит-анализа основаны на внешнем сходстве логистической зависимости с графическим изображением интеграла функции нормального распределения, для которого имеются таблицы, облегчающие вычисления по алгоритму пробит-анализа. Применение пробит-анализа обычно обусловлено также наличием групп из 6—10 подопытных живых объектов, на каждой из которых испытывается определенная доза действующего фактора и учитывается реакция тест-объектов по принципу: погибло—выжило (в процентах). Как будет видно из ниже приводимого примера, последнее условие необязательно, так как можно проводить пробит-анализ с любыми соотношениями результатов реакции тест-объекта, например, выраженными в виде шкалы качественных состояний признака объекта.

Ход вычислений пробит-анализа покажем на примере зависимости зимостойкости интродуцированных в Ленинград 80 экзотических видов деревьев и кустарников от географической широты их естественных ареалов.

Таблица 5.17. Пробит-анализ данных по зимостойкости древесных растений

x^0 с. ш.	y (баллы)	y_1		y_2	y_3	x_2	y_2^2	$x_2 y_2$
1	2	3	4	5	6	7	8	9
20	5	1,0	1,30103	3,09	8,09	0,001	9,55	0,00309
30	4,74	0,948	1,47712	1,63	6,63	0,177	2,66	0,288
35	3,76	0,752	1,54407	0,68	5,68	0,244	0,462	0,166
40	2,86	0,572	1,60206	0,18	5,18	0,302	0,0324	0,0544
45	1,71	0,342	1,65321	-0,41	4,59	0,353	0,168	-0,145
50	1,2	0,24	1,69897	-0,71	4,29	0,398	0,504	-0,283
59	1,01	0,202	1,77085	-0,83	4,17	0,471	0,689	-0,391
				3,63	—	1,946	14,0654	-0,3075

1. В первых двух столбцах табл. 5.17 приведены исходные данные: географическая широта ареалов (в градусах северной широты — x) и зимостойкость (в баллах шкалы Вольфа — y).

2. Выражаем значения функции (y) в долях единицы, для чего на ее максимальное значение (в нашем примере оно равно 5) делим поочередно все остальные значения: $5/5 = 1,0$; $4,74/5 = 0,948$; $3,76/5 = 0,752$ и т. д.; частные от деления записываем в столбце 3.

3. В столбец 4, пользуясь таблицами логарифмов, заносим десятичные логарифмы значений аргумента (x): $\lg 20 = 1,30103$; $\lg 30 = 1,47712$ и т. д. В большинстве случаев достаточно брать при этом лишь три знака мантиссы, особенно, если не требуется высокой точности определения E_{50} .

4. Пользуясь табл. 17П, преобразуем величины y_1 в y_2 . Так, величине 1,0 ($\approx 0,999$) соответствует 3,09, а величине 0,948 соответствует 1,63. Величины y_2 записаны в столбце 5. При повышенных требованиях к точности расчетов можно пользоваться табл. 1 в книге Т. Л. Келли (1966), в которой буква p соответствует нашему обозначению y_1 , а буква x соответствует принятому здесь обозначению y_2 . Если y_1 окажется меньше 0,5, то его значение в той же табл. 1 в книге Келли, следует искать в столбце q , а значение y_2 на той же строке, по-прежнему в столбце, озаглавленном x , но перед y_2 в этом последнем случае ставим знак минус. Для перевода y_1 в y_2 пригодны также и любые другие таблицы, в которых в качестве аргумента берется вероятность, а табулируется в зависимости от нее величина квантилей нормального распределения.

5. Ко всем полученным величинам y_2 прибавляем число 5, и полученные суммы (y_3) заносим в столбец 6: $3,09 + 5 = 8,09$; $1,63 + 5 = 6,63$ и т. д. Число 5, которое прибавляем к y_2 , принимается одинаковым во всех случаях, с его учетом составлены различные таблицы, облегчающие вычисления по некоторым алгоритмам пробит-анализа. После прибавления 5 к величинам y_2 все они становятся положительными, что несколько упрощает вычисления, в чем и заключается смысл этой операции. В терминах теории вероятностей число 5 — это квантиль нормального распределения, которому соот-

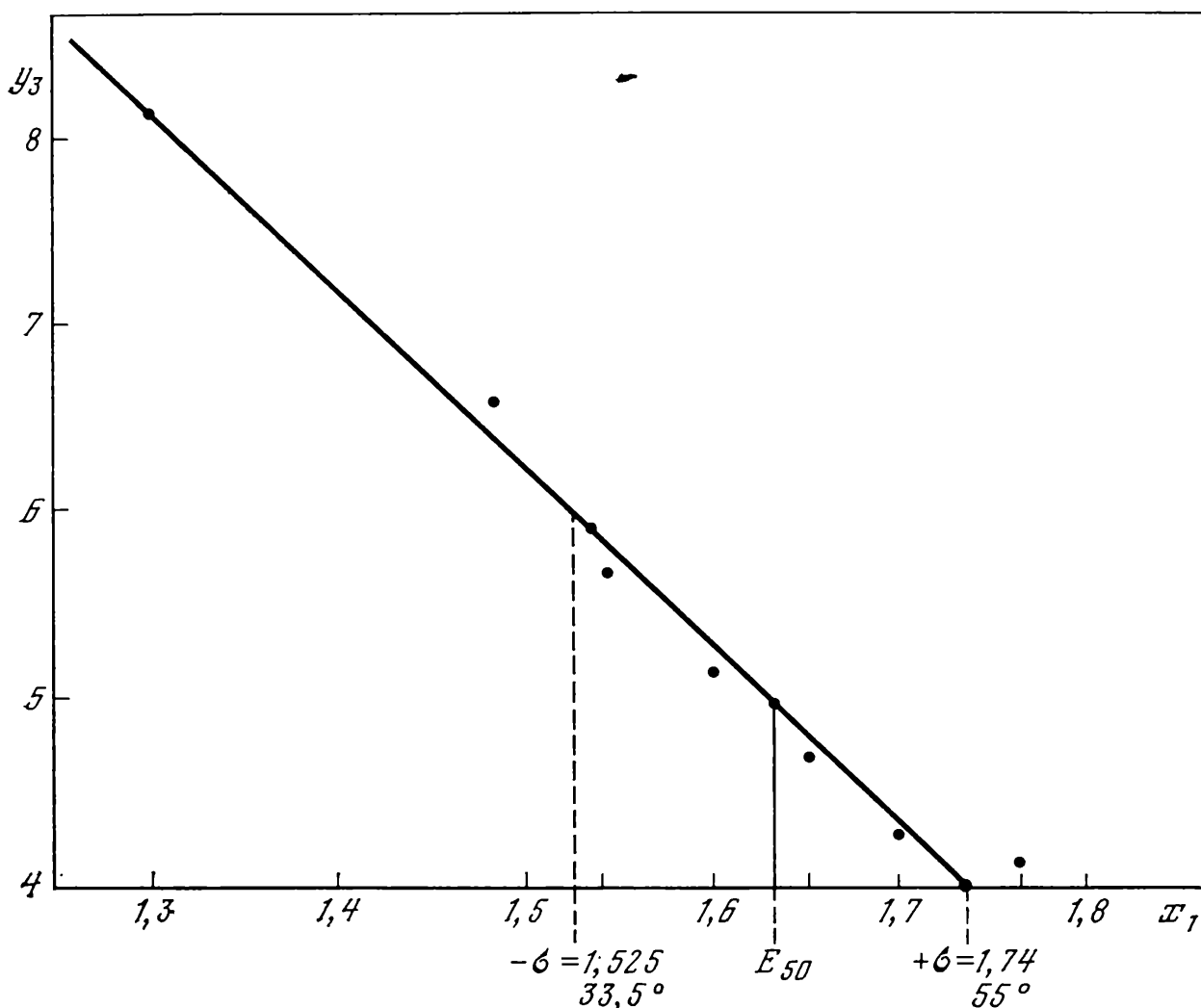


Рис. 34. График пробит-анализа зависимости зимостойкости (y_3) древесных видов растений в Ленинграде от логарифмов широты их естественных ареалов (x_1)

ветствует ничтожная, практически не встречающаяся величина вероятности. Заметим, однако, что теоретически не исключается существование квантиля, равного 5,998, которому соответствует вероятность $1 \cdot 10^{-9}$, при нормальном типе распределения данных. На этом вычисления по алгоритму упрощенного пробит-анализа можно закончить, так как сделано все необходимое для построения графика и определения E_{50} графическим способом.

6. На оси абсцисс откладываем значения x_1 : 1,3; 1,4; и т. д., а на оси ординат — пробиты, т. е. величины y_3 из столбца 6. Строим график точек $x_1 y_3$, который визуальнo с помощью линейки выравниваем прямой линией (рис. 34).

7. На полученной прямой линии находим точку с ординатой 5 и опускаем из нее перпендикуляр на ось абсцисс, на которой у основания этого перпендикуляра читаем значение логарифма величины эффективной дозы. В данном случае $\lg E_{50} = 1,63$, что соответствует северной широте $42,6^\circ$. Эта широта обозначает положение критической линии, разделяющей ареалы зимостойких (севернее $E_{50} = 42,6^\circ$ с. ш.) и незимостойких (южнее $E_{50} = 42,6^\circ$ с. ш.) видов деревьев и кустарников при их интродукции в открытый грунт Ленинграда.

Точности приведенного графического способа определения E_{50} в большинстве случаев достаточно. При повышенных требованиях к точности ее определения и при необходимости проведения проверки расчетов другим способом выполним дальнейшие вычисления, суть которых заключается в определении положения прямой линии (на рис. 34) не визуально, а более точным методом — наименьших квадратов.

1. Из всех величин x_1 (табл. 5.17) для сокращения последующих вычислений вычитаем $A = 1,3$, что представляет собой округленное значение $x_1 = 1,30103$ из первой строки. Полученные разности (x_2) записываем в столбец 7, их сумма: $\Sigma x_2 = 1,946$.

2. Возводим в квадрат величины y_2 из столбца 5 и записываем квадраты в столбец 8: $3,09^2 = 9,55$; $1,63^2 = 2,66$ и т. д., их сумма: $\Sigma y_2^2 = 14,0654$.

3. Перемножаем по строкам числа из столбцов 5 и 7 и результаты заносим в столбец 9: $3,09 \cdot 0,001 = 0,00309$ и т. д., их сумма равна: $\Sigma x_2 y_2 = -0,3075$.

4. Суммируем числа в столбце 5 и вычисляем по формулам:

$$M_1 = A + \frac{\Sigma x_2}{k}, \quad M_2 = \frac{\Sigma y_2}{k},$$

$$\lg E_{50} = M_1 - \frac{\Sigma x_2 y_2 - M_2 \Sigma x_2}{\frac{\Sigma y_2^2}{M_2} - \Sigma y_2}, \quad (5.24)$$

где M_1 — средняя арифметическая логарифмов значений аргумента (x_1) из столбца 4; A — число, вычитаемое из логарифмов аргумента для сокращения вычислений; Σx_2 — сумма остатков, получившихся после вычитания числа A из логарифмов значений аргумента; k — число классов, или строк значений аргумента, в табл. 5.17; M_2 — средняя арифметическая чисел из столбца 5; Σy_2 — сумма чисел столбца 5; $\Sigma x_2 y_2$ — сумма чисел столбца 9; Σy_2^2 — сумма чисел столбца 8.

$$M_1 = 1,3 + \frac{1,946}{7} = 1,578; \quad M_2 = 3,63/7 = 0,5186;$$

$$\lg E_{50} = 1,578 - \frac{-0,3075 - 0,5186 \cdot 1,946}{\frac{14,0654}{0,5186} - 3,63} = 1,634,$$

что соответствует $E_{50} = 42,9^\circ$ с. ш.

При необходимости выделить максимально широкую зону достоверного действия изучаемого эффекта (при уровне значимости 10%), или норму данного явления, следует найти среднее квадратическое отклонение. Наиболее простой и большей частью достаточно точный в данном случае способ — определение сигмы по графику. При этом используется тот факт, что сигма в пробит-анализе расположена на расстоянии 16 и 84% суммарного эффекта действия изучаемого фактора от начальной точки отсчета, что соответствует 4 и 6 пробитам. На оси пробитов — y_3 (см. рис. 34) находим точки 4 и 6, прово-

дим из них перпендикуляры до пересечения с линией графика, откуда в свою очередь опускаем перпендикуляры на ось x_1 , на которой читаем логарифмы: $-\sigma = 1,525$ и $\sigma = 1,74$, что соответствует I и V баллам зимостойкости, или $33,5$ и 55° с. ш. В данном случае, по видимому, эти широты — средние границы умеренной зоны на Европейской части СССР, а $E_{50} = 43^\circ$ — граница между ареалами зимостойких и незимостойких в Ленинграде древесных растений.

§ 5.07. Применение показателей атипичности и аномальности для идентификации особи по комплексу признаков

В тех случаях, когда затруднительно визуальное определение принадлежности отдельной особи к некоторой популяции, сорту, разновидности, форме по качественным признакам можно это сделать методом среднего нормированного отклонения, вычисляемого по комплексу количественных признаков. Считается, что признаки, привлекаемые для этой цели, должны быть независимы друг от друга и значения каждого из них распределены нормально. Большинство признаков растений, находящихся в естественных или близких к тому условиям роста и развития, обычно распределено нормально. Что касается требования о независимости признаков, то оно большей частью невыполнимо, так как все основные, представляющие практический интерес признаки любого организма всегда в той или иной степени взаимосвязаны. Несколько может помочь выполнению условия независимости вычисление корреляционной плеяды данного комплекса признаков и выявление структуры их взаимосвязей с тем, чтобы далее отобрать одинаковые по знаку корреляции и малосвязанные между собой признаки, при том, конечно, условии, если они будут представлять интерес при данном изучении.

Однако не всегда такое сложное статистическое исследование оправдано, имея в виду сравнительно скромную цель, которая при этом будет достигнута. Поэтому указанное условие независимости признаков следует постоянно иметь в виду при биологической интерпретации результатов, а изучение структуры взаимосвязей в комплексе признаков следует проводить лишь тогда, когда это оправдано или большим числом признаков или достаточно высоким уровнем ответственности данного конкретного исследования.

Принадлежность особи к одной из популяций, количественные значения признаков которых заранее определены, обычно устанавливается при помощи метода Хайнке, по сумме квадратов нормированных отклонений. Однако этот метод не дает возможности ответить на следующие важные вопросы: больше или меньше (хуже или лучше) определяемый экземпляр той или иной популяции по комплексу признаков, находится ли данное отклонение в пределах нормы для некоторого комплекса признаков, или выходит за ее пределы, или не относится ни к той, ни к другой популяции. Предлагаемый здесь метод на основе вычисления показателя атипичности позволяет получить ответ на поставленные вопросы и в том числе на основной из

них: об отнесении некоторого экземпляра к одной из совокупностей по комплексу его признаков. Показатель атипичности или среднее нормированное отклонение (5.04), (5.05), вычисляется без учета знаков отклонений

$$\Phi = \frac{\sum |t|}{n}$$

а показатель аномальности — с учетом знаков отклонений:

$$\Phi_1 = \frac{\sum t}{n},$$

где $t = (x - M)/\sigma$ — нормированные отклонения признаков; x — значения признаков определяемой особи; M и σ — средние арифметические и средние квадратические отклонения совокупностей, или популяций, к которым может принадлежать данная особь; n — число признаков в данном комплексе.

Рассмотрим методику применения указанных показателей атипичности на следующем примере. В результате свободного опыления в одной из рядом расположенных посадок сортов гелениума осеннего Катарина и Гранатштерн вырос из самосева экземпляр, обладающий одновременно признаками обоих из указанных сортов. Средние значения и степень варьирования обоих сортов были количественно хорошо известны на основании предыдущих многолетних наблюдений. Заметим, что для целей идентификации особей количественные значения признаков обоих сортов могли бы быть установлены и в данный сезон. В любом случае следует проводить сравнение по каждому признаку по тем его значениям, которые были измерены в одно и то же время вегетационного периода в одной и той же местности и желательно при прочих равных условиях.

В табл. 5.18 приведены конкретные значения четырех признаков у гибридной особи и средние по двум сортам.

1. Вычисляем нормированные отклонения по первому сорту:

$$(150 - 147,2)/12,8 = +0,219; \quad (500 - 503)/279 = -0,0111 \text{ и т. д.}$$

2. Нормированные отклонения по второму сорту:

$$(150 - 151,6)/8,9 = -0,180; \quad (500 - 497)/166 = +0,0181 \text{ и т. д.}$$

3. Суммируем нормированные отклонения без учета и с учетом их знаков и вычисляем показатели атипичности и аномальности для первого:

$$\Phi = 1,057/4 = 0,264; \quad \Phi_1 = 0,239/4 = 0,0597$$

и второго сорта:

$$\Phi = 6,8581/4 = 1,715; \quad \Phi_1 = -2,882/4 = -0,720.$$

Сопоставляя эти показатели по обоим сортам, видим, что гибридный экземпляр по комплексу данных признаков гораздо ближе к сорту Катарина, чем ко второму, так как показатель атипичности пер-

Таблица 5.18. Определение степени близости гибридного экземпляра к его родителям

Признак	Гибрид, x	Сорт Катарина			Сорт Гранатштерн		
		M		t	M	σ	t
Высота растения, см	150	147,2	12,8	+0,219	151,6	8,9	—0,180
Число соцветий	500	503	279	—0,011	497	166	+0,0181
Диаметр соцветий, мм	49	48,1	2,1	+0,429	55,1	1,3	—4,69
Число побегов	15	20,3	13,3	—0,398	9,1	3,0	+1,97
$\sum t $				1,057			6,8581
$\sum t$				+0,239			—2,882
Φ				0,264			1,715
Φ_1				0,0597			—0,720

вого сорта в обеих его модификациях по абсолютной величине меньше, чем у второго сорта: $0,264 < 1,715$ и $|0,0597| < |-0,720|$.

При этом можно сказать, что по своим показателям гибридная особь несколько превышает в целом средние значения для сорта Катарина и сильно уступает в них же сорту Гранатштерн. Оба сорта находятся в пределах нормы по комплексу признаков для данной популяции сортов, так как у них абсолютные величины Φ_1 меньше единицы, причем сорт Гранатштерн по своим признакам находится ниже центра интервала нормы, а сорт Катарина — немного выше его. Однако по своей атипичности сорт Гранатштерн существенно отклоняется от прочей массы сортов ($\Phi = 1,715 > 1,0$), что представляет интерес для внутривидовой систематики гелениума осеннего.

§ 5.08. Критерий пропорциональности в ландшафтной архитектуре

При ландшафтно-архитектурной организации озелеяемого пространства, помимо прочих, во многом интуитивных принципов, обычно учитывают также пропорции объектов. Однако никаких количественных критериев при определении меры пропорции или диспропорции при этом, по крайней мере в настоящее время, не предусматривается. Существующий ныне принцип пропорциональности не имеет выраженной сферы приложения, а понятие «пропорциональное соотношение» не обладает определенным содержанием и практически не отличается от фактов диспропорций в области ландшафтной архитектуры, возникающих из-за произвольных субъективных решений соответствующих проектных ситуаций.

С одной стороны, ландшафтно-архитектурное проектирование — действительно область своеобразного искусства, требующего участия в нем значительной доли интуиции и художественного вкуса. Однако,

с другой стороны, несомненно и то, что под интуицией, как это бывало и раньше во многих недостаточно развитых науках, нередко скрывается некоторая доля нашего познания предмета. Поэтому остановимся на основных принципах пространственного размещения растений, которые, по мнению автора, до сих пор еще не были сформулированы достаточно определенно. Что касается критерия пропорциональности в ландшафтной архитектуре, то он в основном может быть освобожден от влияния интуиции и достаточно четко сформулирован и выражен количественно.

Если попросить ряд людей, обладающих достаточно развитым чувством меры и пропорциональности, определить расстояние, на котором следовало бы посадить дерево заданной высоты от дороги, то в среднем это расстояние достаточно близко совпадало бы у всех участвующих в опыте лиц. Например, дерево высотой 5 м лучше выглядит с расстояния 8 м от дороги. Такой эксперимент при желании нетрудно поставить и найти количественное выражение критерия пропорциональности.

Однако особой необходимости в указанном исследовании нет, так как этот критерий, известный под правилом золотого сечения, эмпирическим путем уже нашли зодчие и скульпторы Древней Греции. Следующий ряд чисел: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, 6765, 10 946, 17 711, 28 657, 46368, 75 025 и т. д. образован по правилу: каждый последующий член равен сумме двух предыдущих. Отношение двух рядом стоящих чисел приведенного ряда, начиная с 5, и считается критерием пропорциональности, или правилом золотого сечения. Это отношение обозначается греческой буквой альфа и равно:

$$\alpha = 1,618, \text{ или } 1/\alpha = 5/8 = 8/13 = 13/21 = 21/34 = 0,618.$$

Число $1/\alpha$ при этом с увеличением номера члена ряда стремится к 0,61804, т. е. является пределом отношения предыдущего к последующему члену.

Последовательность чисел, заданная выше, впервые была установлена математиком средневековья Леонардо из Пизы, более известного под псевдонимом Фибоначчи. Числа Фибоначчи обладают рядом интересных и важных свойств, в силу чего они служат постоянным объектом изучения. В частности, они непосредственно связаны с числами треугольника Паскаля, или биномиальными коэффициентами; отношение соседних чисел Фибоначчи, как говорилось, определяет наиболее гармоничное соотношение размеров. Расположение листьев на стеблях, цветков в соцветиях, а также других органов у видов растений часто следует закону ряда Фибоначчи.

Число Фибоначчи в зависимости от номера его в ряду можно вычислить по формуле Бинэ:

$$u_n = \frac{\left(\frac{1 + \sqrt{5}}{2}\right)^n - \left(\frac{1 - \sqrt{5}}{2}\right)^n}{\sqrt{5}}, \quad (5.25)$$

где u_n — значение n -го члена ряда Фибоначчи, округленного до ближайшего целого числа.

Формула Бино справедлива для любого целого n , которое может быть и со знаком минус, что позволяет продолжить ряд и в отрицательную область. Число α , т. е. предел отношения двух соседних чисел Фибоначчи, равно

$$(1 + \sqrt{5})/2 = 1,6180.$$

Его можно получить также из бесконечной или цепной дроби. Преобразуя бесконечную непрерывную дробь: $1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}$ в виде

$x = 1 + \frac{1}{x}$, найдем ее значение: $x^2 - x - 1 = 0$, откуда $x = (1 \pm \sqrt{5})/2$, положительный корень уравнения равен $\alpha = (1 + \sqrt{5})/2 = 1,6180$.

Тот же результат можно получить из выражения $1/\alpha = \alpha - 1$, которое связывает прямую ($\alpha = 1,61804$) и обратную ($1/\alpha = 0,61804$) формы отношения золотого сечения и далее может быть представлено уравнением $\alpha^2 - \alpha - 1 = 0$, откуда $\alpha = (1 + \sqrt{5})/2$.

Число α связано с числом π через равенство $\pi - \alpha = \pi/6 + 1$, где $\pi - \alpha = 1,5236$, а $\pi/6 = 0,5236$, откуда $\pi = 6(\alpha + 1)/5$, $\alpha = 5/6\pi - 1$, $1/\alpha = 5/6\pi - 2$.

Геометрически золотое сечение представляет собой деление отрезка на две части, большая из которых есть среднее пропорциональное между меньшей частью и длиной всего отрезка (рис. 35, а). Примем длину отрезка AB за единицу, обозначим длину большей части через x , при этом длина меньшей части будет равна $1 - x$ и по условию задачи получим пропорцию

$$\frac{1}{x} = \frac{x}{1 - x},$$

откуда $x^2 + x - 1 = 0$. Положительный корень последнего уравнения равен $(-1 + \sqrt{5})/2$. Следовательно, отношение в последней пропорции равно

$$\frac{1}{-1 + \sqrt{5}} = \frac{2}{-1 + \sqrt{5}} = \frac{1}{2} \sqrt{5},$$

т. е. числу $\alpha = 1,618$.

Деление отрезка AB в среднем и крайнем отношении (как иначе называют золотое сечение) производится следующим образом (рис. 35, б). Длину отрезка AB принимаем за единицу и восстанавливаем перпендикуляр к нему $AE = 1/2$. Гипотенуза образовавшегося треугольника равна

$$EB = \sqrt{1 + \left(\frac{1}{2}\right)^2} = \frac{\sqrt{5}}{2}$$

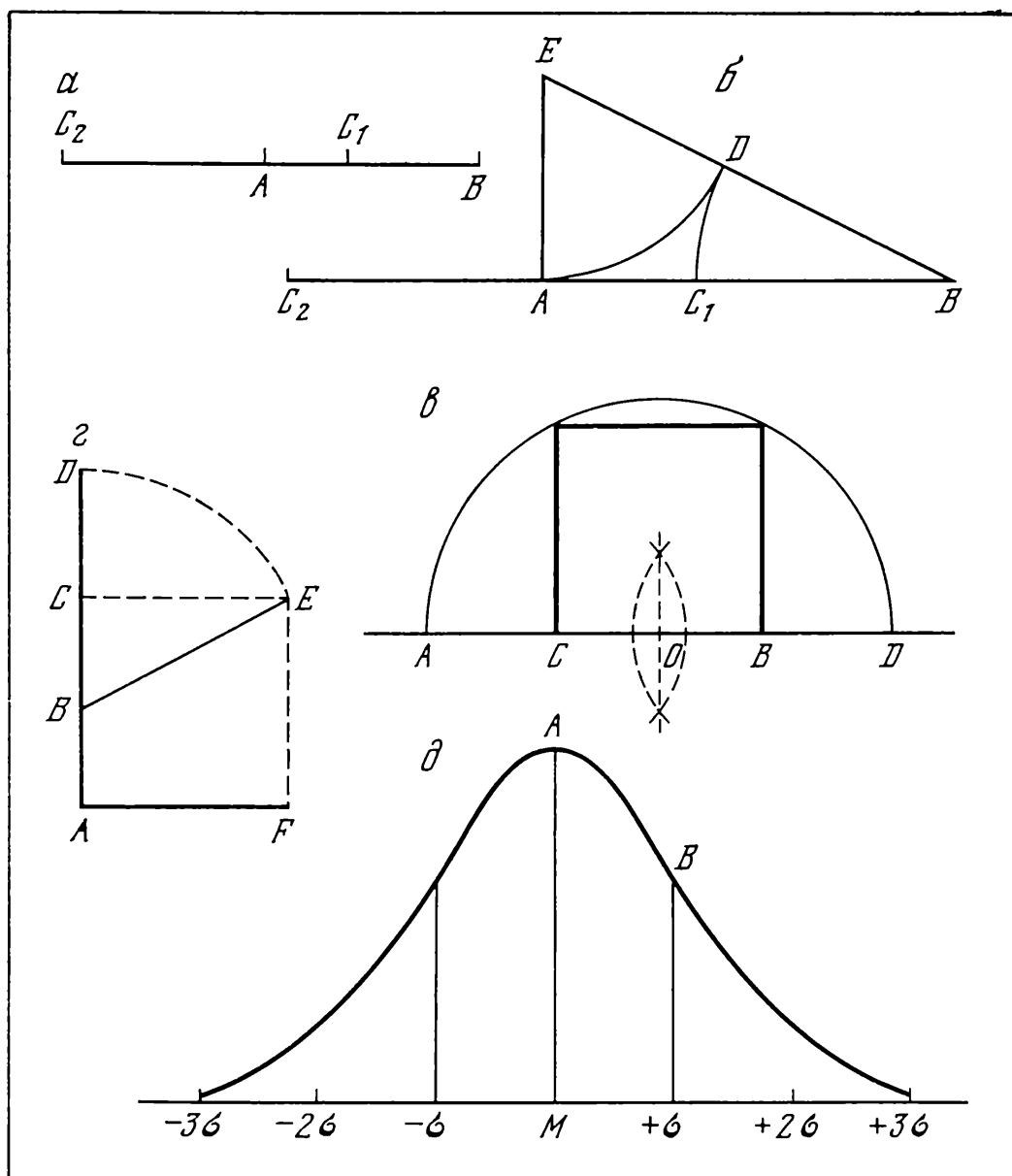


Рис. 35. Графические способы представления отношения золотого сечения

- a* — внутреннее и внешнее деление отрезка AB в отношении золотого сечения: $C_2B/AB = AB/C_1A = \alpha = 1,618$;
- б* — геометрический способ деления отрезка AB в среднем и крайнем отношении: $C_1B/AC_1 = AB/AC_2 = \alpha = 1,618$;
- в* — деление отрезка AD золотым сечением способом построения квадрата, вписанного в полуокружность: $CB/AC = CB/BD = \alpha$;
- г* — построение большего отрезка при заданном меньшем в отношении золотого сечения;
- д* — кривая нормального распределения: максимальная ордината AM относится к ординате $B\sigma$ как $1,618 = \alpha$

Проведя дугу с центром в точке E через точку A , получим точку D . Отрезок

$$BD = (\sqrt{5} - 1)/2 = 1/\alpha.$$

Дуга с центром в точке B , проведенная через точку D , пересекает отрезок AB в точке C_1 , которая делит его в отношении золотого сечения. Откладывая влево отрезок C_2A , равный C_1B , получим точку C_2 внешнего деления золотым сечением отрезка AB .

При построении квадрата, вписанного в полуокружность радиусом AO (рис. 35, *в*), отрезок AB делится точкой C , а также отрезок CD точкой B — в отношении золотого сечения. Большая часть радиуса, разделенного золотым сечением, является стороной правильного десятиугольника, вписанного в данную окружность, что применяется в качестве способа деления окружности на 10 частей. Например, если считать отрезок AB радиусом, то разделив его золотым сечением по способу, показанному на рис. 35, *б*, можно разделить окружность на 10 частей, откладывая по ней циркулем отрезок C_1B . Сторона a правильного десятиугольника, вписанного в окружность радиуса R , равна

$$a = 2R \frac{\sqrt{5} - 1}{4} = \frac{R}{\alpha}$$

Некоторые линейные элементы правильного пятиугольника, в частности его сторона и диагональ, относятся между собой по правилу золотого сечения. В практических построениях встречаются в основном два случая: при заданном меньшем отрезке построить больший отрезок и при заданном большем отрезке найти меньший, находящийся к нему в отношении золотого сечения.

Поэтому приведем еще два, предназначенных для этих целей способа, основу которых составляют уже описанные выше построения. По первому способу находим больший отрезок при заданном меньшем (рис. 35, *г*).

1. Заданный меньший отрезок AC откладываем еще раз под прямым углом к нему, получим отрезок AF .

2. Засечками циркулем из точек C и F находим четвертый угол квадрата — точку E .

3. Отрезок AC делим пополам, находим точку B .

4. Радиусом BE откладываем отрезок BD на продолжении линии AC .

5. Отрезок AD будет относиться к AC как 1 : 0,618, т. е. по правилу золотого сечения.

По второму способу можно найти меньший отрезок, если задан больший, в отношении золотого сечения. Для этого делается то же построение, указанное на рис. 35, *г*, в следующем порядке.

1. Заданный отрезок AC также откладывается еще раз под прямым углом (AF).

2. Находится четвертый угол E квадрата $ACEF$ (засечками циркулем).

3. Радиусом BE откладывается отрезок BD .

4. Отрезок CD есть искомый меньший отрезок, который относится к большему (AC) как 0,618 : 1.

Исследование ряда Фибоначчи до 84 его члена на ЭВМ с целью уточнения числа α показало, что отношения двух стоящих рядом чисел стремятся к $\alpha = 1,61804$ неравномерно, по закону волнообразно затухающей функции.

Нормой различных измеряемых признаков предметов, явлений, объектов служит область центральных вариантов совокупности, распо-

ложенная около средней арифметической (M) и ограниченная с обеих сторон средним квадратическим отклонением ($M \pm \sigma$), против которого находятся точки перегиба кривой нормального распределения, характеризующие моменты качественного изменения явления, в данном случае — моменты перехода типичных вариантов к нетипичным для нормального распределения величин конкретного признака. Поскольку отношение золотого сечения принимается за норму пропорциональности, то не исключено, что оно должно обладать некоторой долей общности со свойствами или параметрами нормального распределения. С целью проверки этого предположения было проведено соответствующее исследование, которое показало, что из всех свойств и параметров нормальной кривой отношение максимальной ординаты к ординате с абсциссами $\pm \sigma$, действительно наиболее близко соответствует понятию нормы пропорциональности. При этом также выявилось, что в понятии нормы в зависимости от степени дробности деления вариант на классы можно различать количественную и качественную форму. Качественным делением совокупности можно считать ее деление на две части: «норма» и «не норма», когда количественно «норма» заключена в интервале $M \pm \sigma$, а «не норма» выходит за пределы этого интервала.

Отношение ординаты, абсцисса которой в правильном нормальном распределении соответствует средней арифметической, медиане и моде, к ординате среднего квадратического отклонения довольно близко соответствующее числу α и служит критерием пропорциональности. Заметим, что равенство ординаты, соответствующей одному нормированному отклонению — числу $5/8$ (т. е. числу $1/\alpha$) от максимальной ординаты, считают одним из критериев нормальности распределения.

Ординату нормальной кривой (z_α), находящуюся в отношении $\alpha = 1,618$ с ее максимальной ординатой (z_0),[§] получаем из выражения:

$$z_\alpha = z_0/\alpha = 0,39894228/1,618 = 0,24656507.$$

Значение z_0 с указанной здесь точностью можно взять, например, из таблиц Т. Л. Келли (1966). Нормированному отклонению 0,98098529 в указанных таблицах соответствует ордината 0,24657117; отношение максимальной ординаты z_0 к ней равно 1,61796, т. е. наиболее близко к числу α . Составив простую пропорцию, уточняем значение нормированного отклонения, соответствующего z_α , оно равно: $t_\alpha = 0,980936$, т. е. несущественно отклоняется от единицы. На рис. 35, δ показаны максимальная ордината (AM) нормального распределения, которая относится к ординате точки перегиба ($B\sigma$) как 1 0,618, т. е. в отношении α -пропорции.

Таким образом, отношение золотого сечения непосредственно связано с нормальным распределением, наиболее общим из типов распределений, известных к настоящему времени и тем самым с понятием нормы для количественных признаков. Поэтому можно считать, что отношение золотого сечения органично и закономерно связано с явлениями, процессами и признаками объектов природы,

Таблица 5.19. Константы, связанные с числом α

Константа	Значение	Константа	Значение	Константа	Значение
α	1,61804	$\alpha/6$	0,26967	$\sqrt{\alpha}$	1,272
$1/\alpha$	0,61804	$\alpha/180$	0,0089891	$\sqrt{2\alpha}$	1,7989
2α	3,2361	$2/\alpha$	1,2361	$\sqrt{\alpha/2}$	0,89947
3α	4,8541	$1/2\alpha = \sin 18^\circ$	0,30901	$\sqrt{1/\alpha}$	0,78615
4α	6,4722	$1/3\alpha$	0,20601	$\sqrt{2/\alpha}$	1,1118
$\alpha/2 = \cos 36^\circ$	0,80902	$1/4\alpha$	0,15451	$1/\alpha^2$	0,38196
$\alpha/3$	0,53935	α^2	2,6181	$\ln \alpha$	0,48122
$\alpha/4$	0,40451	$2\alpha^2$	5,2361	$\ln 1/\alpha$	-0,48122

что может служить дополнительным теоретическим обоснованием для применения числа α в качестве нормы пропорциональности. В табл. 5.19 приведены некоторые величины, применяемые в вычислениях, связанных с числом α .

Линейные размеры объектов, находящихся в отношении золотого сечения, для любых, также и промежуточных между числами Фибоначчи значений, удобно получать с графика (рис. 36), который представляет собой прямую линию, вычерченную по уравнению: $y = 1,61804 x$, или $x = 0,61804 y$, где x — меньший из двух размеров и числитель их отношения; y — больший из размеров, знаменатель отношения x/y .

График на рис. 36 строится следующим образом. После построения координатных осей и разметки их в наиболее часто на практике встречающемся диапазоне измерений на оси абсцисс откладываем 5 см, а на оси ординат — 8 см. Далее строим точку по данным двум координатам и проводим прямую через эту точку и начало координат, после чего график отношений золотого сечения готов.

Искомые размеры, соответствующие отношению 1,618 или обратному отношению 0,618, можно вычислить по приведенным формулам или получить с графика на рис. 36.

Например, дерево высотой $x = 5$ м надо сажать от дороги или другой точки осмотра на расстоянии $y = 1,618 \cdot 5 \approx 8$ м. Травянистый многолетник высотой $x = 40$ см лучше посадить на расстоянии 64 см, что без вычислений можно снять с графика, прочитав это значение против $x = 40$. В необходимых случаях приведенный график можно продолжить в любом (отрицательном или положительном) направлении.

Отношение золотого сечения в ландшафтной архитектуре может применяться во всех случаях, где требуется выдержать гармоничную соразмерность различных элементов озеленительных ансамблей и сопровождающих их форм малой архитектуры. Применение соотношения по золотому сечению между высотой растений и их расстояниями от точки или линии осмотра (рис. 37, а) представляет собой

наиболее распространенный случай в ландшафтной архитектуре. Кроме этого, возможно применение α -пропорции также и во многих других случаях, например при устройстве зеленых заслонов, кулисс, небольших аллей или зеленых стенок, в том случае, если их длина или высота существенно не ограничены другими, например санитарно-гигиеническими или хозяйственными условиями. На рис. 37, б показана кулисса из деревьев ели, длина которой относится к высоте по α -пропорции, что производит более приятное впечатление по сравнению с более удлиненными или

напротив укороченными кулиссами. Аналогично этому отношение высоты трельяжа к его длине как 0,618 : 1 придает более гармоничные пропорции этому устройству (рис. 37, в).

Стрижка зеленой изгороди таким образом, чтобы ее высота относилась к толщине как 1 : 0,618 также повышает ее декоративный эффект (рис. 37, г).

Соотношение по высоте растений между собой в группе может быть во многих случаях построено по принципу α -пропорции (рис. 37, д), что в общем увеличивает декоративность таких композиций.

Конечно, во всех случаях α -пропорцию следует выдерживать лишь в том случае, если нет более важных эстетических, пространственных или других ограничений. Независимо от целей проекта и намеченных средств его реализации решение конкретных ситуаций в ландшафтной архитектуре всегда должно быть комплексным с учетом максимально возможного числа факторов, существенно влияющих на результат, на создание и эксплуатацию данной декоративной композиции.

При размещении заданного числа растений на озеленяемой территории обычно преследуется цель, постоянная почти для всех озеленительных посадок: учитывая другие задачи проекта, максимально удлинить совокупный период цветения композиций. Одновременно с подбором видов и сортов, обеспечивающих максимальную продолжительность цветения композиции, следует подбирать среди них в соответствии с замыслом и критерием пропорциональности растения нужной высоты, а также с желательными окрасками цветков и листвы. В зависимости от того, каким числом видов и сортов располагает озеленитель и какова площадь озеленяемого участка, можно подойти двояко к размещению растений: оно может быть плотным и свободным.

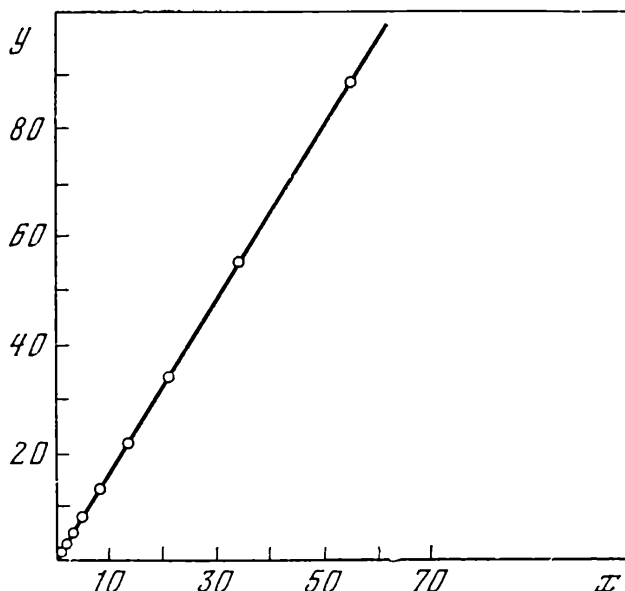


Рис. 36. График критерия пропорциональности по правилу золотого сечения $x/y = 1/\alpha = 0,618$

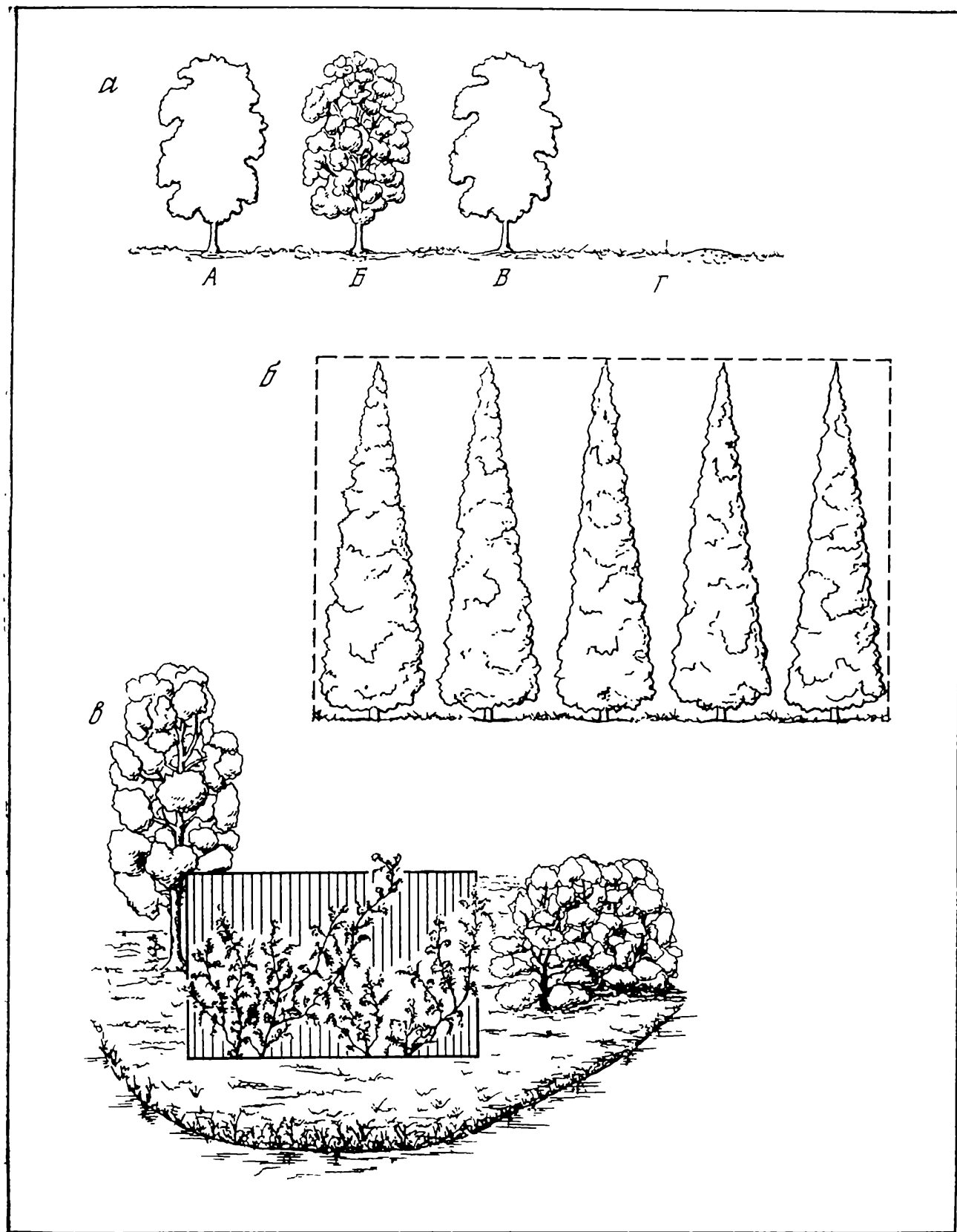
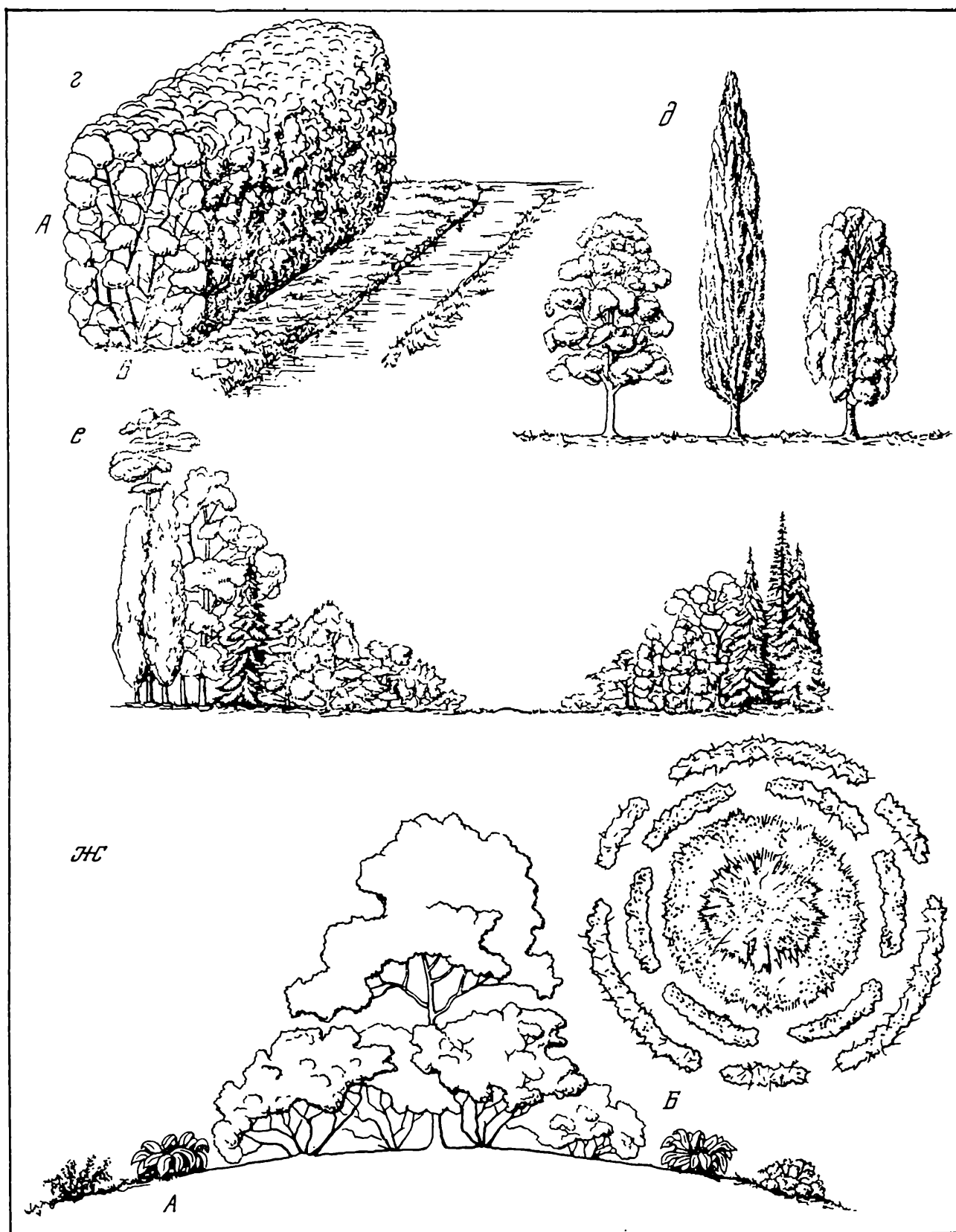


Рис. 37. Применение отношения золотого сечения и приемы компактного размещения растений в ландшафтной архитектуре

а — варианты посадки дерева на разных расстояниях от дорожки; АГ — слишком далеко, БГ — расстояние, при котором высота дерева относится к отрезку БГ, как $5 : 8$, т. е. в отношении золотого сечения, ВГ — дерево посажено слишком близко к дорожке; **б** — рядовая посадка деревьев ели, длина посадки относится к ее высоте, как $1 : 0,618 = \alpha$; **в** — трельяж для вьющихся растений, его длина относится к высоте, как $1 : 0,618 = \alpha$; **г** — высота А



стриженной зеленой изгороди относится к ее ширине Б как $1 : 0,618 = \alpha$; б — соотношение по высоте деревьев в группе по правилу золотого сечения, высота наиболее высокого дерева относится к высоте более низких деревьев, как $1 : 0,618 = \alpha$; в — наиболее компактное размещение на ограниченном по площади участке, обеспечивающее достаточный обзор всех растений данного озеленительного комплекса; ж — размещение в декоративной посадке по принципу зеленой пирамиды, обеспечивающее обзор всех входящих сюда растений: А — вид сбоку, Б — вид сверху, в плане

Наиболее плотное размещение растений на озеленяемой территории при сохранении возможности их осмотра обеспечивается при вогнутом профиле вершин растений в посадках (рис. 37, е). Такое размещение удобно на прямоугольных вытянутых участках, в середине которых проходит линия осмотра (дорога), а по обеим сторонам справа и слева располагают посадки, высота которых постепенно или ступенчато увеличивается по направлению к краям участка. Например, у дороги высевают низкие травы, далее можно — более высокие из них, затем сажают кустарники, при этом располагают их в следующем порядке: низкие, средние, высокие, в таком же по высоте порядке высаживают и деревья.

При иной, например, неправильной округлой или многоугольной форме участка, с максимальной плотностью растения можно расположить на нем по принципу пирамид (рис. 37, ж), которые по территории участка размещаются свободно.

Следует обратить особое внимание на то, что правило золотого сечения лишь дает норму пропорциональности в том или ином случае. Пропорция же представляет собой только одно из средств художественного представления замысла, поэтому ее критерий не может во всех случаях подчиняться правилу золотого сечения. Нередко для усиления или ослабления эффекта массы определенных объектов композиции желательно отклоняться от нормы пропорциональности соответственно в сторону приближения или удаления объекта относительно смотровой линии или точки. Поэтому названный критерий пропорциональности служит именно нормой, благодаря знанию которой становится возможным делать отклонения от нее вполне осознанно, что вносит в ландшафтно-архитектурное проектирование те новые теоретические элементы определенности, место которых занимала интуиция.

* * *

Отметим, что число, подобное «золотому сечению», можно получить из любого целого по формуле

$$\frac{1}{a+x} = x, \quad (5.26)$$

где a — целая часть числа; x — дробная его часть.

Решая уравнение $x^2 + ax - 1 = 0$, которое имеет два корня, можно получить для любого заданного целого a обе формы числа, подобного «золотому сечению», например для $a = 2$:

$$x_1 = \frac{-a - \sqrt{a^2 + 4}}{2} = \frac{-2 - \sqrt{2^2 + 4}}{2} = -2,4142;$$

$$x_2 = \frac{-a + \sqrt{a^2 + 4}}{2} = \frac{-2 + \sqrt{2^2 + 4}}{2} = 0,4142.$$

Для целых чисел от 1 до 20 обе формы (x_1 и x_2) приведены ниже.

	x_1			x_1	
1	0,61805	—1,6181	11	0,0905	—11,091
2	0,41420	—2,4142	12	0,083	—12,083
3	0,30280	—3,3028	13	0,0765	—13,077
4	0,23610	—4,2361	14	0,071	—14,071
5	0,19260	—5,1925	15	0,0665	—15,067
6	0,1622	—6,162	16	0,0625	—16,063
7	0,14005	—7,140	17	0,0585	—17,059
8	0,1232	—8,123	18	0,0555	—18,056
9	0,1097	—9,1095	19	0,0525	—19,053
10	0,099	—10,099	20	0,05	—20,05

§ 5.09. Усреднение и нормирование взвешенных вариационных рядов

При обобщении результатов опытов или наблюдений встречается необходимость усреднить несколько вариационных рядов с тем, чтобы получить одно, но более достоверное в целом эмпирическое распределение некоторого признака. Вариационные ряды, которые надлежит усреднить, должны иметь одну и ту же сетку классовых интервалов, т. е. общие классовой интервал и границы классов; число же классов и объем рядов могут быть различными.

В табл. 5.20 в столбцах 1—5 приведены данные измерений высоты (в см) однолетних растений гелениума осеннего за три года. Требуется вычислить частоты среднего или обобщенного взвешенного вариационного ряда.

Таблица 5.20. Усреднение взвешенных вариационных рядов

Высота растений, см		Годы наблюдений			f_{Σ}	$\frac{f_{\Sigma}}{n}$	f'
Границы класса		1967	1969	1970			
20—31	26	1			1	0,33	0,47
32—43	38	6	1		7	2,33	2,46
44—55	50	21	1	9	31	10,33	9,26
56—67	62	23	11	36	70	23,33	25,39
68—79	74	55	34	58	147	49	50,56
80—91	86	84	61	86	231	77,1	73,16
92—103	98	87	72	69	228	76,1	76,96
104—115	110	54	103	45	202	67,33	58,83
116—127	122	14	39	16	69	23	32,69
128—139	134	13	10	13	36	12	13,19
140—151	146	4	3	5	12	4	3,87
152—163	158		3	5	8	2,67	0,83
164—176	170		1		1	0,33	0,13
$c=12$	$N:$	362	339	342	1043	347,67	347,79 2.
	M	89,8	100,6	91,5		93,9	$k=13$
	σ	21,49	18,79	21,71		21,24	
		24	19	24		23	

В табл. 5.20 частоты за отдельные годы суммированы по строкам, а затем разделены на $n = 3$, т. е. на число усредняемых рядов, независимо от того, сколько частот фактически насчитывается в той или иной строке. Для проверки следует вычислить средние арифметические по отдельным группам (годам), а затем среднюю обобщенного ряда; между ними должно соблюдаться соотношение (1.15):

$$M_{\Sigma} = \frac{89,8 \cdot 362 + 100,6 \cdot 339 + 91,5 \cdot 342}{362 + 339 + 342} = 347,67.$$

Показатели варьирования (σ , ν) усредняемых и обобщенного рядов, как видим из таблицы, также соответствуют изменению средних.

В последнем столбце табл. 5.20 вычислены частоты нормальной кривой, соответствующие показателям обобщенного ряда, которые удовлетворительно совпадают с эмпирическими частотами, что подтверждает предположение о получении более правильной формы эмпирического распределения в результате усреднения рядов, распределения частот которых до этого были более неровными:

Усреднение рядов с одинаковыми интервалами и величинами признака x по осям абсцисс, как видим, несложная процедура. Более трудная задача — получить обобщенный ряд по рядам с неодинаковыми сетками классов по обеим осям и относящихся к различным признакам. Для того чтобы усреднить такие ряды, их необходимо нормировать по одному и тому же правилу. Нормирование рядов по оси ординат — нетрудная задача, для этого достаточно частоты выразить в частостях, т. е. в процентах от объема ряда или в долях единицы. Казалось бы и нормирование величин признаков по оси абсцисс также легко осуществимо после того, как середины классов будут выражены в нормированных отклонениях по формуле $t = (x - M)/\sigma$, однако, как нетрудно убедиться в этом, ряды остаются по-прежнему несравнимыми, ибо почти всегда не совпадают значения середин или границ классов. Таким образом, возникает вопрос, как привести все усредняемые ряды к одинаковой сетке классов по оси абсцисс и в связи с этим привести соответственно также и частоты к новой сетке классов. Иначе говоря, требуется полный перерасчет старой системы координат на новую по обеим осям, который предлагается выполнить следующим образом.

1. В результате решения предыдущего примера в столбце 7 табл. 5.20 были получены частоты обобщенного ряда, варианты которого находятся в столбце 2 той же таблицы. Запишем эти варианты в столбце 1 табл. 5.21, добавив к ним по одному классу в начале и в конце.

2. По серединам классов из столбца 1 табл. 5.21 рассчитаем нормированные отклонения: $t = (x - 93,9)/21,24$ и запишем их в столбце 2. Как видим, диапазон изменения признака стал от $-3,7618$ до $+4,1478$ в нормированных отклонениях. В соответствии с этим размахом варьирования устанавливаем сетку новых классов: от $-3,5$ до $+3,5$ с классовым интервалом $+0,5$, которую запишем в столбце 7. Эта же сетка будет применяться для остальных сравниваемых или

Таблица 5.21. Нормирование взвешенного вариационного ряда

	t	f	f_1	f_3	t_1	t_2	t_3	f_2	$p\%$
1	2	3	4	5	6	7	8	9	10
14	-3,7618	0	0	0,33	-3,7618	-3,5	-3,1968	0,153	0,0396
26	-3,1968	0,33	0,33	2,33	-3,1968	-3	-2,6318	1,03	0,2666
38	-2,6318	2,33	2,33	10,33	-2,6318	-2,5	-2,0669	4,20	1,087
50	-2,0669	10,33	10,33	23,33	-2,0669	-2	-1,5019	11,9	3,080
62	-1,5019	23,33	23,33	49	-1,5019	-1,5	-0,93691	23,4	6,057
74	-0,93691	49	23,33	49	-1,5019	-1	-0,93691	46,1	11,93
86	-0,37194	77,1	49	77,1	-0,93691	-0,5	-0,37194	70,7	18,30
98	0,19303	76,1	77,1	76,1	-0,37194	0	0,19303	76,4	19,77
110	0,758	63,33	76,1	63,33	0,19303	0,5	0,758	69,2	17,91
122	1,323	23	63,33	23	0,758	1	1,323	46,1	11,93
134	1,888	12	23	12	1,323	1,5	1,888	19,5	5,047
146	2,4529	4	12	4	1,888	2	2,4529	10,4	2,692
158	3,0179	2,67	4	2,67	2,4529	2,5	3,0179	3,89	1,007
170	3,5829	0,33	4	2,67	2,4529	3	3,0179	2,71	0,7014
182	4,1478	0	2,67	0,33	3,0179	3,5	3,5829	0,673	0,1742
$k+2=15$	$M=93,9$	347,67	$\sigma=21,24$	$c=12$				386,356	99,9918

усредняемых рядов, поэтому следует предусмотреть достаточную амплитуду величин нормированных отклонений.

3. Найдем ближайшие меньшее (t_1) и большее (t_3) числа из столбца 1 для каждого t_2 . Например, для $-3,5$: $t_1 = -3,7618$, $t_3 = -3,1968$; для -3 : $t_1 = -3,1968$, $t_3 = -2,6318$ и т. д., и запишем их соответственно в столбцах 6 и 8 таблицы.

4. Из столбца 3 выписываем в столбцы 4, 5 частоты f_1 и f_3 , соответствующие значениям t_1 и t_3 . Например, для $t_1 = -3,7618$ величина f_1 равна 0, для $t_3 = -3,1968$ величина f_3 равна 0,33 т.е. эти частоты будут находиться в тех же строках.

5. По формуле

$$f_2 = \frac{(f_3 - f_1)(t_2 - t_1)}{t_3 - t_1} + f_1, \quad (5.27)$$

где f_1 , f_3 — предыдущая и последующая за искомой (f_2) частоты; t_1 , t_3 — предыдущее и последующее нормированные отклонения, или ближайшие меньшее и большее к заданному t_2 , вычисляем частоты f_2 , соответствующие новой классовой сетке (от $-3,5$ до $+3,5$ с интервалом 0,5), которые приведены в столбце 9.

6. Суммируем частоты f_2 , их сумма равна 386,356, что отличается от объема исходного ряда, равного 347,67.

7. Для нормирования полученных частот преобразуем их в доли p , в процентах от их суммы, например,

$$p_1 = \frac{0,153}{386,356} \cdot 100 = 0,0396\%.$$

После этих вычислений нормирование ряда закончено как по оси абсцисс (столбец 7), так и по оси ординат (столбец 10). Отложив величины t_2 по горизонтальной и $p\%$ — по вертикальной осям координат, можно построить график кривой данного эмпирического распределения, который будет сравнимым с графиком кривой распределения любого другого признака. Частоты p , расположенные в одноименных классах t , можно усреднять по любому числу рядов и получать ряды, обобщенные по различным признакам. Изложенный метод был нами применен при изучении явлений нормы и типичности в онтогенезе гелениума осеннего, благодаря чему объективно установлено, что двулетние особи этого растения ближе к естественному и типичному состоянию, чем однолетние и, следовательно, первые более пригодны в качестве тест-объектов для различных опытов, наблюдений и морфологических описаний в систематике этого рода.

§ 5.10. Задачи на составление смесей

В практической и научной деятельности нередко возникает необходимость получить смесь каких-либо компонентов с заданным свойством или оценить результаты смешивания известных компонентов. Рассмотрим на практических примерах подобные задачи и способы их решения.

1. Какое октановое число будет у смеси, состоящей из 30 л бензина следующих марок: А66—5 л, А72 — 10 л, А76 — 10 л, А93 — 5 л.

Принцип решения этой и всех подобных задач на составление смесей заключается в применении способа взвешенного вариационного ряда, который в данном случае выглядит так:

Варианта	Вес
66	5
72	10
76	10
93	5
<hr/>	
С у м м а:	30

Если вычислить среднюю арифметическую этого ряда (1.13), то это и будет ответом на поставленную задачу:

$$M = \frac{66 \cdot 5 + 72 \cdot 10 + 76 \cdot 10 + 93 \cdot 5}{5 + 10 + 10 + 5} = 75,8,$$

т. е. октановое число данной смеси равно 75,8 или А76. Число компонентов, т. е. число членов подобного ряда, может быть любым, а промежутки (интервалы) между вариантами могут быть равными и неравными между собой, варианты могут быть отрицательными, на способ вычисления средней все это не влияет.

2. Механизатор располагает 10 л бензина марки А72, сколько к ним следует добавить бензина А93, чтобы получилась смесь с октановым числом А76?

Составляем вариационный ряд:

72	10
93	x
<hr/>	
С у м м а:	10

Средняя этого двучленного ряда уже известна, она равна 76. Пишем выражение подобно тому, которое вычисляли в первой задаче:

$$76 = \frac{72 \cdot 10 + 93 \cdot x}{10 + x},$$

где x — неизвестное число литров бензина А93, которое отсюда равно:

$$x = \frac{10(76 - 72)}{93 - 76} = 2,35 \text{ л.}$$

3. Имеется 10 л бензина А72. Какой марки бензин к ним следует добавить, чтобы получилось 20 л бензина с октановым числом А76? Способ решения этой задачи тот же, что и для двух первых: составляем вариационный ряд, затем пишем выражение для средней арифметической, откуда приемами школьной алгебры находим неизвест-

ную величину.

$$\begin{array}{r} 72 \quad 10 \\ x \quad 10 \\ \hline \text{Сумма: } 20 \end{array} \quad 76 = \frac{72 \cdot 10 + x \cdot 10}{20},$$

$$\text{откуда } x = \frac{20 \cdot 76 - 72 \cdot 10}{10} = \frac{10(2 \cdot 76 - 72)}{10} = 80,$$

т. е. следует добавить 10 л бензина А80.

4. Сколько дистиллированной воды требуется добавить к 0,5 л серной кислоты плотностью 1,84, чтобы получить электролит плотностью 1,27? Плотность воды, как известно, равна 1, поэтому возможно составить вариационный ряд:

$$\begin{array}{r} 1,84 \quad 0,5 \\ 1 \end{array}$$

Средняя этого ряда равна:

$$1,27 = \frac{1,84 \cdot 0,5 + 1 \cdot x}{0,5 + x},$$

$$\text{откуда } x = \frac{0,5(1,84 - 1,27)}{1,27 - 1} = 1,05 \text{ л},$$

т. е. к указанному количеству кислоты следует добавить 1,05 л воды, чтобы получился электролит с плотностью 1,27.

5. Сколько воды (x) требуется добавить к смеси четырех растворов соли: 2% — 1 л, 5% — 0,5 л, 3% — 2 л, 4% — 1 л, чтобы получился раствор с концентрацией 2,5%? Составляем вариационный ряд, пятый компонент которого представляет собой неизвестное количество раствора с нулевой концентрацией:

$$\begin{array}{r} 2\% \quad 1 \\ 3\% \quad 2 \\ 4\% \quad 1 \\ 5\% \quad 0,5 \\ 0\% \quad x \\ \hline \end{array}$$

Средняя арифметическая этого ряда равна:

$$2,5 = \frac{2 \cdot 1 + 3 \cdot 2 + 4 \cdot 1 + 5 \cdot 0,5 + 0 \cdot x}{1 + 2 + 1 + 0,5 + x}$$

откуда

$$x = \frac{2 \cdot 1 + 3 \cdot 2 + 4 \cdot 1 + 5 \cdot 0,5 + 0 \cdot x - 2,5(1 + 2 + 1 + 0,5)}{2,5} = 1,3 \text{ л}.$$

6. В воды реки сбрасывают производственные отходы (в т в день) 5 предприятий (А — Д) со следующей концентрацией вредных веществ:

Предприятие	Концентрация вредных веществ	Количество отходов
А	2	20
Б	2,5	15
В	2,7	12
Г	3	18
Д	3,1	7
Сумма:		72

Надо сравнить эти предприятия между собой по степени загрязнения ими окружающей среды. Для этого получим произведения концентраций на количество отходов по каждому предприятию: $2 \cdot 20 = 40$; $2,5 \cdot 15 = 37,5$; $32,4$; 54 ; $21,7$, их сумма: $185,6$, средняя арифметическая равна $185,6 / 72 = 2,58$. Концентрация отходов предприятия *Б* всего ближе к общей средней, следовательно, оно типично в отношении загрязнения данных вод. Произведение концентрации на количество отходов наименьшее ($21,7$) у предприятия *Д*, которое, таким образом, меньше других загрязняет воды; наиболее сильно загрязняет воды предприятие *Г*, у которого указанное произведение равно 54 .

§ 5.11. Определение неприступных расстояний

В различных экспедиционных и стационарных условиях встречается необходимость определения размеров или расстояний, непосредственное измерение которых невозможно или сильно затруднено. В подобных случаях применяют различные способы косвенных измерений, связанные с несложными расчетами. Многие из них известны и применяются еще с древности, что свидетельствует об их постоянном практическом значении.

а. Радиус круглого объекта. По этому способу можно определить поперечный размер озера или впадины, ямы, основания холма и любых других объектов, очертание которых в плане близко к окружности или полуокружности. Определим радиус окружности объекта, центр которой находится в точке *О* (рис. 38). Наблюдатель из точки *А* фиксирует крайние точки *В* и *М*, находящиеся на окружности, затем измеряет угол *ВАМ*, делит его пополам и находит точку *Д*, которую можно найти также, разделив пополам дугу *ВМ*; не измеряя угла *ВАМ* при условии, что данная дуга доступна для измерения. Проведя через точку *А* линию, перпендикулярную к *АД*, наблюдатель измеряет угол α , при этом: $\alpha + \beta = 90^\circ$. Измерив расстояние $AD = h$, по формуле

$$R = \frac{h \cos \alpha}{1 - \cos \alpha}, \quad (5.28)$$

можно вычислить радиус объекта. Например, при $h = 30$ м, $\alpha = 62^\circ$ $R = 30 \cdot 0,4695 / (1 - 0,4695) = 26,5$ м, т. е. диаметр окружности равен 53 м. Диаметр на местности можно определить также через длину окружности или полуокружности, если они доступны измерению.

б. Неприступное расстояние на плоскости, например ширину реки, каньона, ущелья, оврага, впадины, можно определить, если рядом с объектом имеется свободный участок, достаточный для геометрических построений (рис. 39). Для этого на местности строят два прямоугольных треугольника *АВС* и *АСД* с общей стороной *АС* и двумя углами *ВАС* и *САД*, равными друг другу, при помощи вехи в точке *М*. Расстояние *ВС* равно *CD*, измерив последнее на местности, узнаем искомое расстояние *BC*.

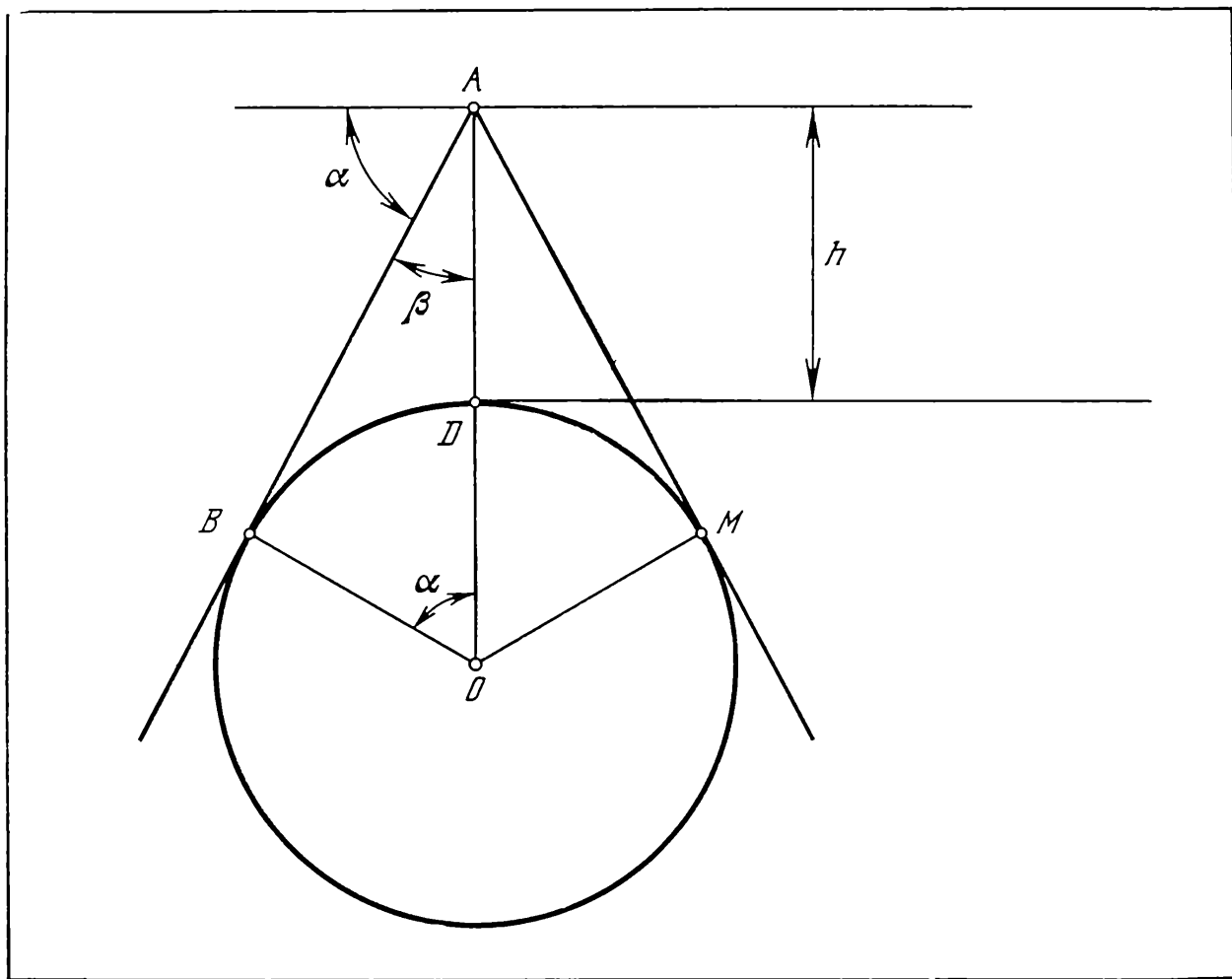


Рис. 38. Определение недоступного для измерения радиуса окружности

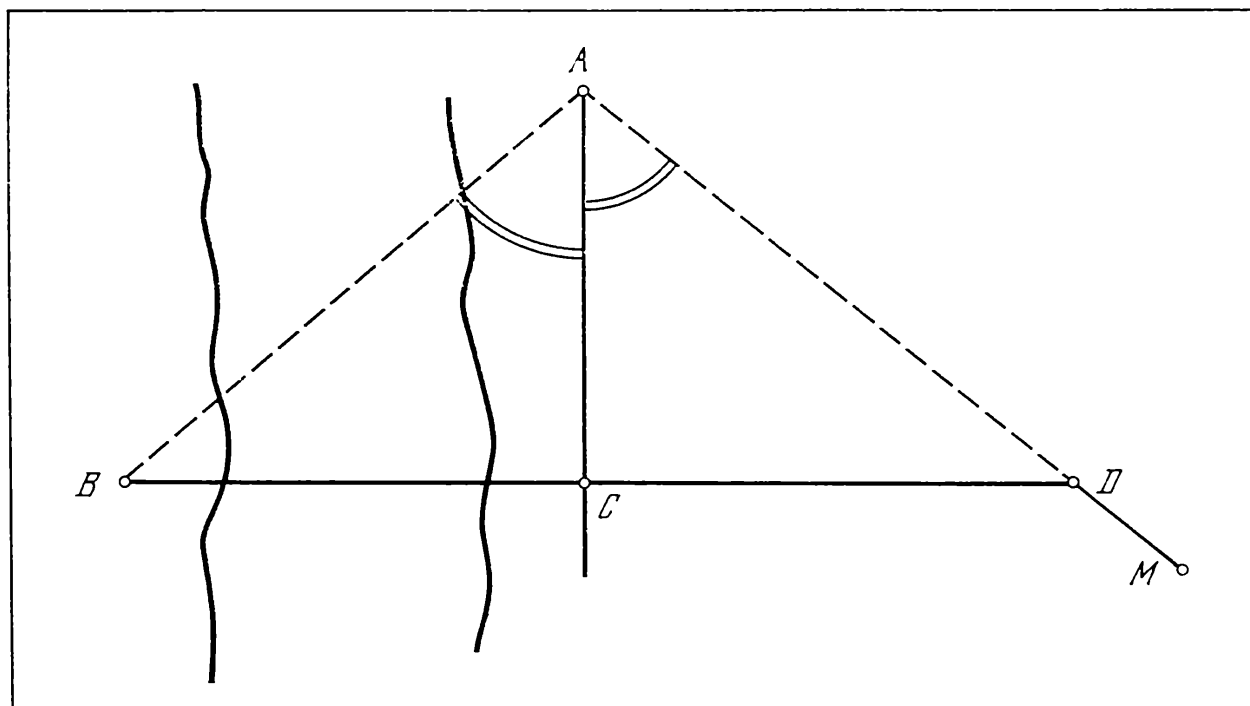


Рис. 39. Определение неприступного расстояния BC

горизонтали $АС$. При этом следует учитывать высоту глаз наблюдателя, когда производится визирование из точек A и B в точку D для определения углов α и β . Точность описанных способов сильно зависит от правильного определения углов, поэтому последние следует считать по возможности точнее.

г. Удобны на практике измерения неприступных расстояний с помощью несложного приспособления, представляющего собой две планки или стержни, закрепленные под прямым углом так, что длинный стержень (qn) разделен коротким (pm) в отношении $qm/mn = 1 \div 0,5$. Большая часть (qm) длинного стержня при этом равна короткому стержню (pm), с которым она образует равносторонний прямой угол. С помощью отвеса длинную часть (qn) приспособления устанавливаем вертикально, как показано на рис. 41, 1 и визируем из точки B так, чтобы точки p , q и A оказались на одной линии. Высота дерева равна: $H = BC + a$. Если расстояние BC из-за рельефа местности измерить трудно, то на месте первого визирования делают отметку, поворачивают прибор короткой частью длинного стержня вверх и производят второе визирование из точки D . Высота объекта будет равна: $H = DB + a$.

Размеры приспособления — высотомера могут быть различными: чем они больше, тем точность определения будет выше. Минимальные размеры: длина большого стержня — 150 мм, короткого — 100 мм.

д. На рис. 41, 2 дана схема определения высоты при помощи другого приспособления — квадрата из фанеры, картона, оргстекла, размером не менее 250 × 250 мм. На одном из углов прикреплен отвес, длина которого чуть больше диагонали квадрата. Производится визирование из произвольной точки на точку A вдоль одной из сторон квадрата. Высота дерева на схеме равна:

$$H = \frac{Kp \cdot qn}{pq} + a. \quad (5.30)$$

Сторону квадрата, по которой двигается грузик отвеса, рекомендуется разметить миллиметровыми делениями. При расчете по формуле (5.30) сначала следует получить отношение qn/pq , на которое умножить длину отрезка Kp , затем прибавить к нему отрезок a — высоту точки визирования. Точность определения высоты по этому способу также возрастает при увеличении размера квадрата — высотомера.!

е. На рис. 41, 3 приведена схема приблизительного определения высоты при помощи записной книжки и карандаша, который вставляется в корешок переплета. Производится визирование из произвольной точки так, чтобы угол книжки p , конец карандаша q и вершина дерева A были на одной линии. Высота дерева равна:

$$H = \frac{pK \cdot qn}{pn} + a. \quad (5.31)$$

Следует при визировании удерживать книжку в вертикальном положении, желательно с помощью отвеса.

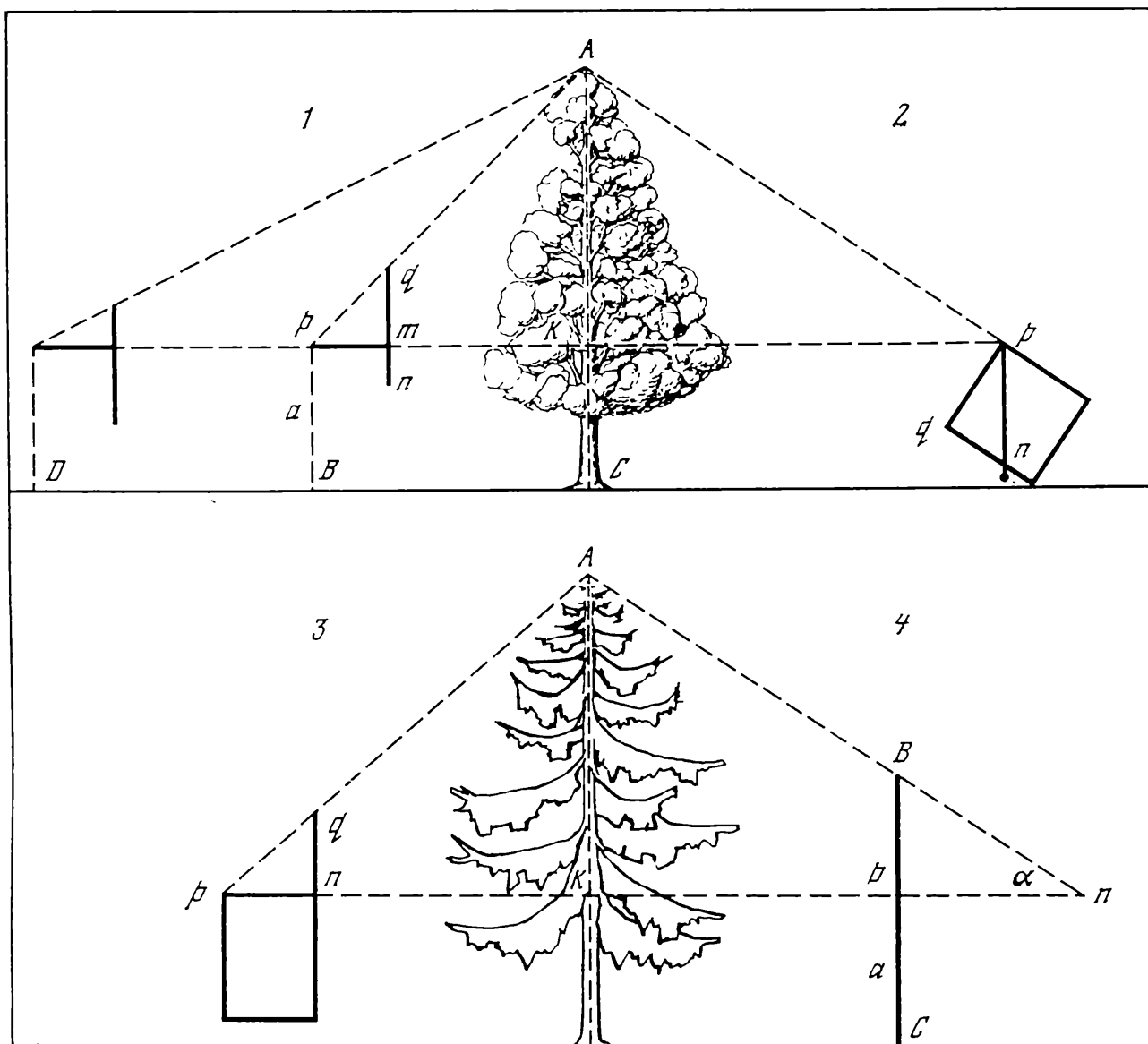


Рис. 41. Измерение высоты объектов при помощи двухречного высотомера (1), квадрата с отвесом (2), записной книжки и карандаша (3) и шеста (4)

ж. На рис. 41, 4 дается схема измерения высоты объекта при помощи шеста BC , который посредством отвеса установлен вертикально в произвольной точке. Определяемая высота равна:

$$H = \frac{nK \cdot Bp}{np} + a. \quad (5.32)$$

Если имеется хотя бы простой прибор для измерения углов, то высоту объекта можно определить по формуле

$$H = nK \cdot \operatorname{tg} \alpha + a, \quad (5.33)$$

что позволит обойтись без шеста.

* *

Значения тригонометрических функций тех или иных углов лучше определять по таблицам, однако при крайней необходимости приближенные значения всех этих функций можно определить при помощи несложного графического приема, точность результатов которого

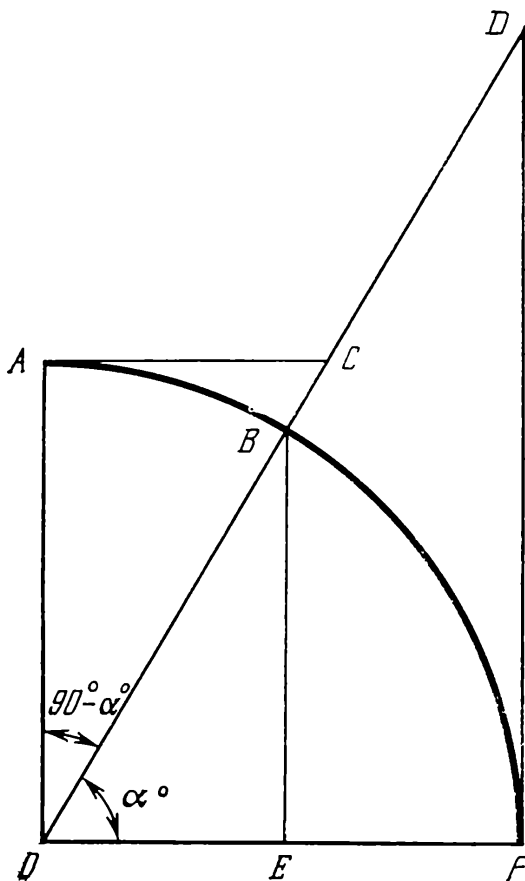


Рис. 42. Графическое определение величины тригонометрических функций

OA — радиус, линии: EB — синуса, OE — косинуса, DF — тангенса, AC — котангенса, OD — секанса, OC — косеканса

сильно зависит от аккуратности геометрических построений. Измерив на рис. 42 или на местности величины соответствующих отрезков, находим значения тригонометрических функций по формулам:

$$\begin{aligned}\sin \alpha &= \frac{BE}{R}, & \cos \alpha &= \frac{OE}{R} \\ \operatorname{tg} \alpha &= \frac{DF}{R}, & \operatorname{ctg} \alpha &= \frac{AC}{R}, \\ \sec \alpha &= \frac{OD}{R}, & \operatorname{cosec} \alpha &= \frac{OC}{R},\end{aligned}\quad (5.34)$$

где $R = OA = OB = OF$.

Заметим, что последние четыре функции, кроме формул известных из литературы, связаны уравнением

$$(\operatorname{cosec} \alpha - \sec \alpha)^2 = (\operatorname{tg} \alpha - 1)^2 + (\operatorname{ctg} \alpha - 1)^2. \quad (5.35)$$

Для построения прямых углов на местности можно использовать приспособление, известное с древних времен. Нетолстую веревку длиной 12 м, которая должна как можно меньше вытягиваться при ее натяжении, следует петлями, или узлами, или металлическими кольцами разделить на три части: 3, 4 и 5 м. Можно взять любые единицы измерения длины веревки, важно лишь, чтобы соблюдалось соотношение 3 4 5. Натянув с равномерным небольшим усилием веревку по колышкам, укрепленным в грунте на границах указанных отрезков, получим треугольник, в котором стороны 3 и 4 образуют угол 90° , стороны 3 и 5 — 60° , стороны 4 и 5 — 30° .

§ 5.12. Формулы перевода результатов измерений температуры и других градусных мер

Перевод величин измерений температуры из одной шкалы в другую можно сделать по формулам:

1. $C^\circ = \left(\frac{9}{5} n + 32 \right)^\circ F,$
2. $C^\circ = \left(\frac{4}{5} n \right)^\circ R,$
3. $C^\circ = (n + 273,15)^\circ K,$
4. $F^\circ = \left[\frac{5}{9} (n - 32) \right]^\circ C,$
5. $F^\circ = \left[\frac{4}{9} (n - 32) \right]^\circ R,$
6. $F^\circ = \left[\frac{5}{9} (n - 32) + 273,15 \right]^\circ K,$
7. $R^\circ = \left(\frac{5}{4} n \right)^\circ C,$
8. $R^\circ = \left(\frac{9}{4} n + 32 \right)^\circ F,$

$$\begin{aligned}
9. \quad R^{\circ} &= \left(\frac{5}{4} n + 273,15 \right)^{\circ} \text{ К}, & 11. \quad K^{\circ} &= \left[\frac{9}{5} (n - 273,15) + 32 \right]^{\circ} \text{ F}, \\
10. \quad K^{\circ} &= (n - 273,15)^{\circ} \text{ C}, & 12. \quad K^{\circ} &= \frac{4}{5} (n - 273,15)^{\circ} \text{ R},
\end{aligned}
\tag{5.36}$$

где C° , F° , R° , K° — соответственно величины температуры по шкалам Цельсия, Фаренгейта, Реомюра и Кельвина.

Например, 4° R по Цельсию равны по формуле (5.36, 7):

$$R^{\circ} = \frac{5}{4} \cdot 4 = 5^{\circ} \text{ C}.$$

В различных биологических исследованиях, в частности связанных с изучением циркадных и других ритмов физиологических процессов, фотопериода, фенологических явлений, нередко требуется переводить результаты измерений в дуговых мерах в единицы времени и обратно. При этом можно руководствоваться следующими соотношениями:

- 1 час времени = 15° дуги;
 - 1 минута времени = $0^{\circ}15'$ дуги;
 - 1 секунда времени = $0^{\circ}15''$ дуги (табл. 32П);
 - 1 час времени = 0,041667 суток;
 - 1 минута времени = 0,000694 суток;
 - 1 секунда времени = 0,000012 суток (табл. 33П);
 - 1° дуги = 4 минуты времени;
 - 1 минута дуги = 4 секунды времени;
 - 1 секунда дуги = 0,07 секунды времени (табл. 34П).
- Тропический (календарный) год = 365,2432 средних солнечных суток.

§ 5.13. Определение площади сосудов на срезе стебля

При анатомических исследованиях нередко встречается необходимость определить долю (процент) площади, занятую объектами круглой формы, расположенными в свою очередь на фоне круга, например процент площади сосудов на срезе стебля, которую можно найти по формуле:

$$p = 0,25 \frac{\sum d^2}{r^2}, \tag{5.37}$$

где p — доля площади сосудов; d — диаметр сосудов; r — радиус среза или любого круглого фона.

Диаметры сосудов измеряют на всей площади круга (среза). Если измерения диаметров сделаны на половине площади среза, то

$$p = 0,5 \frac{\sum d^2}{r^2}, \tag{5.38}$$

где обозначения те же. Например, на всей площади среза радиусом 5 мм насчитали 5 сосудов определенного типа с диаметрами 0,2; 0,2; 0,3; 0,4; 0,4 мм; сумма их квадратов: $\sum d^2 = 0,49 \text{ мм}^2$, а доля площади: $p = 0,5 \frac{0,49}{25} = 0,0098$, т. е. 0,98% от общей площади среза.

§ 5.14. Комплексная оценка надежности результатов массовых фенологических наблюдений

От степени надежности исходного фактического материала зависит многое: сама целесообразность его обработки, выбор методов для последней и ценность получаемых далее выводов. Вероятность получения надежных результатов наблюдений даже при аккуратной их регистрации обычно снижается за счет влияния отдельных методологических экзогенных (внешних) и эндогенных (неотъемлемо присущих объекту наблюдений) источников ошибок и варьирования, которые при желании могут быть идентифицированы по отдельности. Однако при обобщении и научной обработке результатов каких-либо массовых учетов или наблюдений нередко возникает вопрос, насколько надежны такие данные не по отдельным факторам, а в целом, при этом наибольший интерес для исследователя представляет единый комплексный показатель надежности, который позволил бы обоснованно обобщить результаты действия основных источников ошибок, влияющих на надежность данного массива фактических данных.

Таким образом, задача заключается в том, как можно оценить общую или комплексную надежность наблюдений, если известна вероятность, с которой ожидается получить достоверный результат по каждому отдельному источнику ошибок из числа влияющих на результаты конкретных наблюдений.

Для примера оценим комплексную надежность фенологических наблюдений, которые постоянно проводятся в ботанических садах обычно по одному и тому же контингенту видов растений. На достоверность регистрируемых наблюдателем в журнале фенологических дат по всему массиву наблюдаемых видов оказывают влияние, в частности, следующие факторы: 1) интервал наблюдений; 2) правильность определения систематической принадлежности наблюдаемых растений; 3) субъективизм при оценке фенофаз; 4) различная наследственность особей в пределах вида; 5) микроразности почвы, рельефа, освещенности, температуры и других экологических факторов.

Предположим, что 1) фенонаблюдения проводятся через 5 дней; 2) около 20% наблюдаемых видов вызывают сомнение в правильности их определения; 3) при оценке фенофаз допускается около 10% ошибок, вызванных индивидуальными особенностями наблюдателей; 4) около 15% недостоверности данных вызвано тем, что наблюдалось лишь по одному экземпляру каждого вида, вследствие чего не учитывалось влияние различной наследственности особей в пределах вида; 5) влияние всех микроразностей экологических факторов составляет около 20% отклонений фенодат.

Наблюдается, например, совокупность, состоящая из 100 видов деревьев и кустарников, и предполагается, что у всех наблюдаемых видов имеется равная возможность быть зафиксированными наблюдателем при изменении фенофаз в один из пяти дней интервала. При этом только $1/5$ всех дат, или 20%, может быть зафиксирована наблюдателем верно, так как в остальные четыре дня наблюдений не проводится. На оставшиеся 20%, или 20 дат, накладывается доля

недостоверности, вызываемая неточностями определения видов, равная 20%, что составляет от 20 дат — 4 даты. Следовательно, доля надежных данных снижается до $20 - 4 = 16$ дат. Субъективные ошибки в оценке фенофазы составляют 10% от 16 дат, что равно 1,6 даты, откуда доля надежных данных снижается до $16 - 1,6 = 14,4$ даты. 15% от 14,4 равно 2,16 и $14,4 - 2,16 = 10,8$, что соответствует числу надежных дат, оставшемуся после учета влияния разной наследственности особей в пределах вида. И наконец, учитывая влияние микроразностей экологических факторов, получаем: 80% от 10,8 дат, что составляет 8,64 даты.

Отметим, что факторы, принятые во внимание при исчислении комплексной надежности, должны быть независимыми друг от друга, т. е. несвязанными друг с другом значимыми коррелятивными связями.

Таким образом, учитывая все ошибки, заданные выше предположительными их величинами, мы получаем, что из 100 регистрируемых дат заслуживают доверия 8,64% фенодат.

Особенно большая ошибка в фенодаты вносится интервалом наблюдений; при интервале в 5 дней 80% фенонаблюдений можно поставить под сомнение, при интервале через 7 дней лишь 14,3% получаемых фенодат заслуживает доверия ($100/7 = 14,3$).

Представив вышеуказанные проценты надежных фенодат в виде вероятностей, т. е. в виде отношения числа фактических событий к числу возможных, получим: $20/100 = 0,2$; $80/100 = 0,8$; $90/100 = 0,9$; $75/100 = 0,75$; $80/100 = 0,8$.

Согласно теореме умножения вероятностей получим следующую вероятность определения надежных фенодат, или комплексную оценку надежности массива наблюдений:

$$P_0 = 0,2 \cdot 0,8 \cdot 0,9 \cdot 0,75 \cdot 0,8 = 0,0864.$$

Вычтя это число из единицы, определим вероятность получения ненадежных данных:

$$P = 1 - P_0 = 1 - 0,0864 = 0,9136. \quad (5.39)$$

Способы повышения общей надежности массива данных можно установить путем анализа отдельных влияющих на них источников ошибок.

Частота посещений наблюдаемых растений (так же как и многих других объектов) почти полностью задается наблюдателем. Поэтому влияние этого фактора может быть сведено к любой малой величине при ежедневных или более частых наблюдениях. Однако ежедневные наблюдения при большом числе объектов становятся весьма обременительными. Поэтому целесообразно применять меняющийся интервал, если известны примерные сроки наступления интересующих фенофаз, или имеются признаки, позволяющие предвидеть приблизительно массовое наступление фенофаз. В эти периоды интервал наблюдений сокращается до минимума, а в периоды с небольшой частотой фенологических явлений интервал может быть значительно увеличен. При ежегодных наблюдениях, например в дендрариях,

ботанических садах, обычно примерно известны периоды, на которые падает большинство фиксируемых фенодат. На эти периоды целесообразно запланировать ежедневное посещение объектов. Вне периодов с учащенным наступлением фенодат интервал наблюдений можно увеличить. При наблюдениях изменений любых признаков состояния растений необходимо определять интервал наблюдений, руководствуясь в основном скоростью изменения признака во времени и изменчивостью данного признака у наблюдаемого объекта. Чем выше скорость изменения признака и чем больше его изменчивость, тем короче должен быть интервал наблюдений.

Снижение надежности наблюдений, вызываемое неточностями определения систематической принадлежности наблюдаемых растений, в принципе также может быть устранено. Тем не менее в силу постоянных изменений видового состава растений в ботанических садах и по другим причинам практически не существует массива наблюдаемых растений, в котором все названия полностью соответствуют действительности. Элиминировать этот вид недостоверности материала можно лишь постепенно, в течение ряда лет, широко используя сбор и хранение гербария для сомнительных видов.

Влияние субъективизма при оценке фенофаз, различной наследственности особей в пределах вида, микроразностей почвы, климата и тому подобных факторов может быть оценено постановкой специальных опытов с применением дисперсионного анализа их результатов, что позволит выявить значимость влияния таких факторов на изменение фенодат.

На надежность получаемых фенодат влияют также факторы: число лет наблюдений и число наблюдаемых растений. Степень влияния первого фактора выясняется лишь после математической обработки данных. Если после некоторого числа лет наблюдений средняя арифметическая некоторой фенодаты окажется недостоверной (например, по критерию Стьюдента), то наблюдения следует продолжить до такого числа лет, которое будет достаточным для достоверности средней.

При наблюдениях одного экземпляра по каждому виду нет никакой уверенности, что именно он не окажется нетипичным или даже аномальным для данного вида. Поэтому фенологические наблюдения следовало бы проводить на таком числе фиксированных экземпляров каждого вида, которое обеспечивало достоверность получаемой средней арифметической по каждой из наблюдаемых фенофаз.

Вариабельность фенодат по разным фенофазам, как правило, различается. Например, коэффициент вариации по фенофазе — начало осеннего опадания листьев у фанерофитов в Ленинграде составляет в среднем 5,7%, а по фенофазе конец цветения — 30,2%. Поэтому желательное число лет наблюдений по каждой фенофазе окажется различным. Весенние и осенние фенофазы более сильно зависят от климата, поэтому при выборе фенофазы, по фенодатам которой определяется число лет наблюдений, следует отдать предпочтение летним фенофазам, на которые меньше влияют климатические факторы. Более других в этом отношении подходит фенофаза начало цветения,

коэффициент вариации фенотат по которой для фанерофитов равен в среднем 25%.

Массив регистрируемых наблюдений не будет пригоден вовсе для научной обработки и какого-либо практического применения в том случае, если он не отличается от случайного набора чисел, не содержащего полезной информации о данном явлении. Для оценки степени отклонения данного массива наблюдений от случайного набора чисел можно руководствоваться эмпирически установленным критерием, по которому комплексная надежность данных должна быть не менее 10%, или $P_0 = 0,1$. При таком уровне надежности ожидается получение не менее 10% достоверных данных от общего их числа.

В зависимости от числа наблюдений, доверительной вероятности и от числовой величины комплексной надежности, последняя может быть найдена с различной точностью, что естественно отразится на решении вопроса о положении P_0 относительно принятого ее критического значения (10%).

Оценить точность полученной величины комплексной надежности данных как эмпирической частоты можно при помощи построения для нее доверительного интервала по следующему способу.

1. Проверяем, выполняются ли одновременно два условия:

а) $NP_0 > 4$ и б) $N(1 - P_0) > 4$, что необходимо для применения далее следующих формул, основанных на законе нормального распределения.

2. Находим границы доверительного интервала по формуле:

$$J^{\pm} = \frac{P_0 + t \left(\frac{0,5t}{N} \pm \sqrt{\frac{1}{N} \left[P_0(1 - P_0) + \frac{0,25t^2}{N} \right]} \right)}{1 + \frac{t^2}{N}}, \quad (5.40)$$

где J^{\pm} — границы доверительного интервала; P_0 — величина комплексной надежности; t — нормированное отклонение, соответствующее принятой доверительной вероятности; N — объем массива данных.

Ввиду приблизительного характера подобных расчетов рекомендуется оценивать достоверность величины P_0 , начиная с доверительной вероятности $P(0,90)$, для которой $t = 1,6$, и тогда формулу (5.40) можно конкретизировать следующим образом:

$$J^{\pm} = \frac{0,782NP_0 + 1 \pm \sqrt{1,56NP_0(1 - P_0) + 1}}{0,782N + 2}, \quad (5.41)$$

где обозначения те же.

При увеличении N формула (5.40) в пределе принимает вид:

$$J^{\pm} = P_0 \pm t \sqrt{\frac{P_0(1 - P_0)}{N}}, \quad (5.42)$$

где обозначения те же.

Формулой (5.42) можно пользоваться лишь при одновременном выполнении трех условий: а) $N \geq 200$; б) $NP_0 > 10$; в) $N(1 - P_0) > 10$.

Решая в зависимости от конкретных данных одно из выражений — (5.40), (5.41) или (5.42) — сначала с минусом, а затем с плюсом перед знаком радикала, получаем соответственно нижнюю и верхнюю границы доверительного интервала.

3. Если окажется, что $NP_0 < 4$, $N(1 - P_0) < 4$, то доверительный интервал для P_0 следует определить по таблице доверительного интервала для параметра биномиального распределения [например, (Янко, 1961, табл. 28, с. 79, 181)]. В названной таблице обозначения соответствуют следующим из принятых здесь: $x = NP_0$, $n = N$.

Приведем пример расчета границ доверительного интервала по всем трем формулам и по таблице.

При заданных выше исходных данных: $N = 100$, $P_0 = 0,0864$, примем доверительную вероятность $P(0,90)$, для которой $t = 1,6$.

Проверяем условие нормальности:

$$NP_0 = 100 \cdot 0,0864 = 8,64; NP_0 > 4;$$

$$N(1 - P_0) = 100 \cdot 0,9136 = 91,36; N(1 - P_0) > 4.$$

Следовательно, можно применить формулу (5.40):

$$J_1^\pm = \frac{0,0864 + 1,6 \left(\frac{0,5 \cdot 1,6}{100} \pm \sqrt{\frac{1}{100} \left[0,0864(1 - 0,0864) + \frac{0,25 \cdot 1,6^2}{100} \right]} \right)}{1 + \frac{1,6^2}{100}},$$

откуда $J_1^- = 0,0512$; $J_1^+ = 0,142$.

По формуле (5.41) можно проверить этот результат:

$$J_2^\pm = \frac{0,782 \cdot 100 \cdot 0,0864 + 1 \pm \sqrt{1,56 \cdot 100 \cdot 0,0864(1 - 0,0864) + 1}}{0,782 \cdot 100 + 2},$$

откуда $J_2^- = 0,0512$; $J_2^+ = 0,142$.

Формулой (5.42) при указанных исходных данных пользоваться нельзя, так как: а) $N < 200$; б) $NP_0 < 10$. Однако выполним расчет для тех же данных с иллюстративной целью. По формуле (5.42):

$$J_3^\pm = 0,0864 \pm 1,6 \sqrt{\frac{0,0864(1 - 0,0864)}{100}},$$

откуда $J_3^- = 0,0414$; $J_3^+ = 0,131$.

Так как эти границы доверительного интервала существенно отличаются от полученных по формуле (5.40), рекомендуется постоянно выполнять условия, указанные для формулы (5.42). По таблице для параметра биномиального распределения (Янко, 1961) при $x = 100 \cdot 0,0864 \approx 9$, $n = 100$, $n - x = 91$ (принимая ближайшее большее значение $n - x = 100$), при доверительной вероятности 0,95 (в таблице Янко приводятся интервалы только для двух вероятностей: 0,95 и 0,99) находим: $J_4^- = 0,038$; $J_4^+ = 0,151$. В качестве границ доверительного интервала при заданной вероятности 0,90 принимаем результат, полученный по формуле (5.40).

Таким образом, комплексная надежность $P_0 = 0,0864$ массива из 100 фенонаблюдений находится в интервале: $0,0512 < P_0 < 0,142$, т. е. в данном массиве наблюдений содержится не менее 5% и не более 14% полезной информации, пригодной для научной обработки.

Остальная часть наблюдений, от 86 до 95%, представляет собой массив мало отличающийся от случайного набора чисел. Поскольку нижняя граница найденного доверительного интервала меньше принятого критического значения: $5\% < 10\%$ делаем вывод, что данный массив фенонаблюдений не может служить надежным исходным материалом для дальнейших исследований.

Ввиду приблизительного характера большинства подобных расчетов границы доверительного интервала для величины комплексной надежности в случаях, не требующих особой точности, можно не вычислять, ограничиваясь указанным выше условием: $P_0 \geq 10\%$.

ПРИЛОЖЕНИЯ¹

Таблица 1П. Число классов в зависимости от объема выборки

Объем выборки N	Число классов k	Объем выборки N	Число классов k
40—60	6—8	100—200	9—12
60—100	7—10	200—500	12—17
		и больше	

Таблица 2П. Значения критериев ϑ_N и ϑ_1 на уровнях значимости 0,01 (верхнее число) и 0,05 (нижнее число)*

N	$\vartheta_{N(1)}$	N	$\vartheta_{N(1)}$	N	$\vartheta_{N(1)}$
4	0,991 0,955	13	0,520 0,410	22	0,414 0,320
5	0,916 0,807	14	0,502 0,395	23	0,407 0,314
6	0,805 0,689	15	0,486 0,381	24	0,400 0,309
7	0,740 0,610	16	0,472 0,369	25	0,394 0,304
8	0,683 0,554	17	0,460 0,359	26	0,389 0,299
9	0,635 0,512	18	0,449 0,349	27	0,383 0,295
10	0,597 0,477	19	0,439 0,341	28	0,378 0,291
11	0,566 0,450	20	0,430 0,334	29	0,374 0,287
12	0,541 0,428	21	0,421 0,327	30	0,369 0,283

* Если вычисленные значения ϑ_N или ϑ_1 превышают табличные, то варианта отбрасывается

¹ Таблицы приложения, кроме 23П и 24П, заимствованы из источников, указанных в списке литературы.

Таблица 2Па. Критические значения разности между двумя крайними вариантами совокупности

Объем совокупности, N	Уровни значимости				Объем совокупности, N	Уровни значимости			
	0,10	0,05	0,01	0,005		0,10	0,05	0,01	0,005
2	2,33	2,77	3,64	3,97	80	0,83	1,04	1,50	1,70
3	1,79	2,17	2,90	3,20	90	0,82	1,03	1,50	1,67
10	1,18	1,46	2,03	2,27	100	0,81	1,02	1,47	1,57
20	1,03	1,27	1,80	2,00	200	0,75	0,95	1,38	1,55
30	0,96	1,20	1,70	1,90	300	0,72	0,91	1,32	1,50
40	0,91	1,15	1,63	1,85	400	0,70	0,88	1,30	1,47
50	0,88	1,11	1,60	1,80	500	0,68	0,87	1,28	1,45
60	0,86	1,08	1,57	1,76	1000	0,65	0,83	1,22	1,40
70	0,84	1,06	1,53	1,72					

Таблица 3П. Значения критерия Стьюдента t

Число степеней свободы, ν	Доверительные уровни			Число степеней свободы, ν	Доверительные уровни		
	$P_1=95\%$	$P_2=99\%$	$P_3=99,9\%$		$P_1=95\%$	$P_2=99\%$	$P_3=99,9\%$
	$P_1=0,95$	$P_2=0,999$	$P_3=0,999$		$P_1=0,95$	$P_2=0,99$	$P_3=0,999$
1	12,706	63,657	—	24	2,064	2,797	3,745
2	4,303	9,925	31,598	25	2,060	2,787	3,725
3	3,182	5,841	12,941	26	2,056	2,779	3,707
4	2,776	4,604	8,610	27	2,052	2,771	3,690
5	2,571	4,032	6,859	28	2,048	2,763	3,674
6	2,447	3,707	5,959	29	2,045	2,756	3,659
7	2,365	3,499	5,405	30	2,042	2,750	3,646
8	2,306	3,355	5,041	35	2,030	2,724	3,591
9	2,262	3,250	4,781	40	2,021	2,704	3,551
10	2,228	3,169	4,587	45	2,014	2,690	3,520
11	2,201	3,106	4,437	50	2,008	2,678	3,496
12	2,179	3,055	4,318	55	2,004	2,669	3,476
13	2,160	3,012	4,221	60	2,000	2,660	3,460
14	2,145	2,977	4,140	70	1,994	2,648	3,435
15	2,131	2,947	4,073	80	1,989	2,638	3,416
16	2,120	2,921	4,015	90	1,986	2,631	3,402
17	2,110	2,898	3,965	100	1,982	2,625	3,390
18	2,101	2,878	3,922	120	1,980	2,617	3,373
19	2,093	2,861	3,883	>120	1,960	2,5758	3,2905
20	2,086	2,845	3,850				
21	2,080	2,831	3,819		$IV_1=5\%$	$IV_2=1\%$	$IV_3=0,1\%$
22	2,074	2,819	3,792		или	или	или
23	2,069	2,807	3,767		$IV'_1=0,05$	$IV'_2=0,01$	$IV'_3=0,001$

Таблица 4П. Значения коэффициента h для приближенного определения среднего квадратического отклонения

Объем выборки, N	h	Объем выборки, N	h	Объем выборки, N	h
2	1,1	6	2,5	15—30	4,0
3	1,1	7	2,7	30—200	5,0
4	2,1	9	3,0	200—1000	6,0
5	2,3	15	3,5	>1000	7,0

Таблица 5П. Значение функции $\psi(t) = \frac{0,39894}{e^{0,5t^2}}$ (перед числами в таблице надо поставить 0,)

t	0	1	2	3	4	5	6	7	8	9
0,0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29430	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08938	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
4,0	00013	00009	00006	00004	00002	00002	00001	00001	00000	00000

Таблица 6П. Арифметический треугольник

											Коэффициенты долей частных групп при $\rho=0,5$					
					1		1									
$n=2$								1			0,25					
$n=3$				1			3		1		0,125					
$n=4$					1			4		1	0,0625					
$n=5$						10		10			1	0,03125				
	6		1	6		15		15		1		0,015625				
$n=7$		1			21		35		21		1	0,0078125				
$n=8$			1	8			70		56		28	8	1	0,00390625		
$n=9$		1			36	84		126		84	36		9	1	0,001953125	
$n=10$	1		10		45	120		210		210	120	45		10	1	0,0009765625

Примечание. Таблицу 6П можно легко продолжить до любого n , складывая два рядом стоящих числа строки и записывая сумму в следующей строке в промежутке между ее слагаемыми.

Таблица 7П. Минимальные значения показателя корреляции рангов при данном числе вариант

Число ва- риант N	Значение ρ на двух уровнях значимо- сти *		Число ва- риант N	Значение ρ на двух уровнях значимо- сти *		Число ва- риант N	Значение ρ на двух уровнях значимо- сти *	
	$W'_1=0,05$	$W'_2=0,01$		$W'_1=0,05$	$W'_2=0,01$		$W'_1=0,05$	$W'_2=0,01$
4	0,80	—	10	0,55	0,73	16	0,43	0,58
5	0,80	0,90	11	0,53	0,70	17	0,41	0,56
6	0,77	0,89	12	0,50	0,67	18	0,40	0,55
7	0,68	0,86	13	0,48	0,64	19	0,39	0,53
8	0,60	0,81	14	0,46	0,62	20	0,38	0,52
9	0,58	0,77	15	0,44	0,60			

* ρ недостоверно, если оно меньше табличного.

Таблица 8П. Значение выражений $N(N-1)$; $\frac{2}{N(N-1)}$ и ошибка коэффициента корреляции рангов Кендэла

N	$N(N-1)$	$\frac{2}{N(N-1)}$		N	$N(N-1)$	$\frac{2}{N(N-1)}$	
2	2	1	—	17	272	0,00735	0,1786
3	6	0,3333	—	18	306	0,00654	0,1726
4	12	0,1667	—	19	342	0,00585	0,1671
5	20	0,1000	—	20	380	0,00526	0,1622
6	30	0,0667	—	21	420	0,00476	0,1577
7	42	0,0476	—	22	462	0,00433	0,1535
8	56	0,0357	0,2886	23	506	0,00395	0,1496
9	72	0,0278	0,2665	24	552	0,00362	0,1460
10	90	0,0222	0,2483	25	600	0,00333	0,1426
11	110	0,0182	0,2337	26	650	0,00308	0,1396
12	132	0,0152	0,2214	27	702	0,00285	0,1367
13	156	0,0128	0,2099	28	756	0,00264	0,1338
14	182	0,0110	0,2007	29	812	0,00246	0,1312
15	210	0,0095	0,1922	30	870	0,00230	0,1289
16	240	0,0083	0,1847	31	930	0,00215	0,1265

Таблица 9П. Значения критерия F' при различных доверительных уровнях

$\nu(2)$ $\nu(1)$	1	2	3	4	5	6	8	12	16	24	50	∞
При $P_1 = 95\%$												
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	246,5	249,0	251,8	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,43	19,45	19,47	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,69	8,64	8,58	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,84	5,77	5,70	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,60	4,53	4,44	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,92	3,84	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,49	3,41	3,32	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,20	3,12	3,03	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,98	2,90	2,80	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,82	2,74	2,64	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,70	2,61	2,50	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,60	2,50	2,40	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,51	2,42	2,32	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,44	2,35	2,24	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,39	2,29	2,18	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,33	2,24	2,13	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,29	2,19	2,08	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,25	2,15	2,04	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,21	2,11	2,00	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,18	2,08	1,96	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,15	2,05	1,93	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,13	2,03	1,91	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,11	2,00	1,88	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	2,09	1,98	1,86	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	2,07	1,96	1,84	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	2,05	1,95	1,82	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	2,03	1,93	1,80	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	2,02	1,91	1,78	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	2,00	1,90	1,77	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,99	1,89	1,76	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,94	1,83	1,70	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,90	1,79	1,66	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,87	1,76	1,63	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,85	1,74	1,60	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,81	1,70	1,56	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,79	1,67	1,53	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,77	1,65	1,51	1,32
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,76	1,64	1,49	1,30
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,75	1,63	1,48	1,28
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,72	1,60	1,45	1,25
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,71	1,59	1,44	1,22
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,69	1,57	1,42	1,19
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,68	1,55	1,39	1,15
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,67	1,54	1,38	1,13
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,66	1,54	1,38	1,11
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,65	1,53	1,36	1,08
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,64	1,52	1,35	1,00

Таблица 9П. (Продолжение)

$v^{(1)}$ $v^{(2)}$	1	2	3	4	5	6	8	12	16	24	50	∞
При $P_2 = 99\%$												
1	4052	4999	5403	5625	5764	5859	5981	6106	6169	6234	6302	6366
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,42	99,44	99,46	99,48	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,83	26,60	26,35	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	14,15	13,93	13,69	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,29	9,89	9,68	9,47	9,24	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,52	7,31	7,09	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,27	6,07	5,85	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,48	5,28	5,06	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,92	4,73	4,51	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,52	4,33	4,12	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,21	4,02	3,80	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,98	3,78	3,56	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,78	3,59	3,37	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,62	3,43	3,21	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,48	3,29	3,07	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,37	3,18	2,96	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,27	3,08	2,86	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,20	3,00	2,79	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	3,12	2,92	2,70	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	3,05	2,86	2,63	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,99	2,80	2,58	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,94	2,75	2,53	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,89	2,70	2,48	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,85	2,66	2,44	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,81	2,62	2,40	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,78	2,58	2,36	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,74	2,55	2,33	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,71	2,52	2,30	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,68	2,49	2,27	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,66	2,47	2,24	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,07	2,74	2,56	2,37	2,13	1,90
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,48	2,29	2,05	1,80
45	7,23	5,11	4,25	3,77	3,45	3,23	2,94	2,61	2,43	2,23	1,99	1,75
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,38	2,18	1,94	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,32	2,12	1,87	1,60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,78	2,45	2,28	2,07	1,82	1,53
80	6,96	4,88	4,04	3,56	3,26	3,04	2,74	2,42	2,24	2,03	1,78	1,49
90	6,92	4,85	4,01	3,53	3,23	3,01	2,72	2,39	2,21	2,00	1,75	1,45
100	6,90	4,82	3,98	3,51	3,21	2,99	2,69	2,37	2,19	1,98	1,73	1,43
125	6,84	4,78	3,94	3,47	3,17	2,95	2,66	2,33	2,15	1,94	1,69	1,37
150	6,81	4,75	3,91	3,45	3,14	2,92	2,63	2,31	2,13	1,92	1,66	1,33
200	6,76	4,71	3,88	3,41	3,11	2,89	2,60	2,28	2,09	1,88	1,62	1,28
300	6,72	4,68	3,85	3,38	3,08	2,86	2,57	2,24	2,06	1,85	1,59	1,22
400	6,70	4,66	3,83	3,37	3,06	2,85	2,56	2,23	2,04	1,84	1,57	1,19
500	6,69	4,65	3,82	3,36	3,05	2,84	2,55	2,22	2,03	1,83	1,56	1,16
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,53	2,20	2,01	1,81	1,54	1,11
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,99	1,79	1,52	1,00

Таблица 9П. (Окончание)

$v(2)$ $v(1)$	1	2	3	4	5	6	8	12	24	∞
При $P_3 = 99,9\%$										
1	405284	500000	540379	562500	576405	585937	598144	610667	623497	636619
2	998,5	999,0	999,2	999,2	999,3	999,3	999,4	999,4	999,5	999,5
3	167,5	148,5	141,1	137,1	134,6	132,8	130,6	128,3	125,9	123,5
4	74,14	61,25	56,18	53,44	51,71	50,53	49,00	47,41	45,77	44,05
5	47,04	36,61	33,20	31,09	29,75	28,84	27,64	26,42	25,14	23,78
6	35,51	27,00	23,70	21,90	20,81	20,03	19,03	17,99	16,89	15,75
7	29,22	21,69	18,77	17,19	16,21	15,52	14,63	13,71	12,73	11,69
8	25,42	18,49	15,83	14,39	13,49	12,86	12,04	11,19	10,30	9,34
9	22,86	16,39	13,90	12,56	11,71	11,13	10,37	9,57	8,72	7,81
10	21,04	14,91	12,55	11,28	10,48	9,92	9,20	8,45	7,64	6,76
11	19,69	13,81	11,56	10,35	9,58	9,05	8,35	7,63	6,85	6,00
12	18,64	12,97	10,80	9,63	8,89	8,38	7,71	7,00	6,25	5,42
13	17,81	12,31	10,21	9,07	8,35	7,86	7,21	6,52	5,78	4,97
14	17,14	11,78	9,73	8,62	7,92	7,43	6,80	6,13	5,41	4,60
15	16,59	11,34	9,34	8,25	7,57	7,09	6,47	5,81	5,10	4,31
16	16,12	10,97	9,00	7,94	7,27	6,81	6,19	5,55	4,85	4,06
17	15,72	10,66	8,73	7,68	7,02	6,56	5,96	5,32	4,63	3,85
18	15,38	10,39	8,49	7,46	6,81	6,35	5,76	5,13	4,45	3,67
19	15,08	10,16	8,28	7,26	6,61	6,18	5,59	4,97	4,29	3,52
20	14,82	9,95	8,10	7,10	6,46	6,02	5,44	4,82	4,15	3,38
21	14,59	9,77	7,94	6,95	6,32	5,88	5,31	4,70	4,03	3,26
22	14,38	9,61	7,80	6,81	6,19	5,76	5,19	4,58	3,92	3,15
23	14,19	9,47	7,67	6,69	6,08	5,65	5,09	4,48	3,82	3,05
24	14,03	9,34	7,55	6,59	5,98	5,55	4,99	4,39	3,74	2,97
25	13,88	9,22	7,45	6,49	5,88	5,46	4,91	4,31	3,66	2,89
26	13,74	9,12	7,36	6,41	5,80	5,38	4,83	4,24	3,59	2,82
27	13,61	9,02	7,27	6,33	5,73	5,31	4,76	4,17	3,52	2,75
28	13,50	8,93	7,19	6,25	5,66	5,24	4,69	4,11	3,46	2,70
29	13,39	8,85	7,12	6,19	5,59	5,18	4,64	4,05	3,41	2,64
30	13,29	8,77	7,05	6,12	5,53	5,12	4,58	4,00	3,36	2,59
40	12,61	8,25	6,60	5,70	5,13	4,73	4,21	3,64	3,01	2,23
60	11,97	7,76	6,17	5,31	4,76	4,37	3,87	3,31	2,69	1,90
120	11,38	7,31	5,79	4,95	4,42	4,04	3,55	3,02	2,40	1,56
∞	10,83	6,91	5,42	4,62	4,10	3,74	3,27	2,74	2,13	1,00

Таблица 10П. Значения критерия хи-квадрат

Число степеней свободы, ν	Уровни значимости		
	$W'_1 = 0,05$	$W'_2 = 0,01$	$W'_3 = 0,001$
1	3,841	6,635	10,827
2	5,991	9,2100	13,815
3	7,815	11,345	16,268
4	9,488	13,277	18,465
5	11,070	15,086	20,517
6	12,592	16,812	22,457
7	14,067	18,475	24,322
8	15,507	20,090	26,125
9	16,919	21,666	27,877
10	18,307	23,209	29,588
11	19,675	24,725	31,264
12	21,026	26,217	32,909
13	22,362	27,688	34,528
14	23,685	29,141	36,123
15	24,996	30,578	37,697
16	26,296	32,000	39,252
17	27,587	33,409	40,790
18	28,869	34,805	42,312
19	30,144	36,191	43,820
20	31,410	37,566	45,315
21	32,671	38,932	46,797
22	33,924	40,289	48,268
23	35,172	41,638	49,728
24	36,415	42,980	51,179
25	37,652	44,314	52,620
26	38,885	45,642	54,052
27	40,113	46,963	55,476
28	41,337	48,278	56,893
29	42,557	49,588	58,302
30	43,773	50,892	59,703

Таблица 10Па. Критические значения распределения хи-квадрат

Число с тепе- ней свобо- ды, ν	Уровни значимости					Число степе- ней свобо- ды, ν	Уровни значимости				
	0,3	0,1	0,05	0,01	0,001		0,3	0,1	0,05	0,01	0,001
1	2	3	4	5	6	1	2	3	4	5	6
1	1,07	2,71	3,84	6,63	10,8	51	55,8	64,3	68,7	77,4	88,0
2	2,41	4,61	5,99	9,21	13,8	52	56,8	65,4	69,8	78,6	89,3
3	3,67	6,25	7,81	11,3	16,3	53	57,9	66,5	71,0	79,8	90,6
4	4,88	7,78	9,49	13,3	18,5	54	58,9	67,7	72,2	81,1	91,9
5	6,06	9,24	11,1	15,1	20,5	55	60,0	68,8	73,3	82,3	93,2
6	7,23	10,6	12,6	16,8	22,5	56	61,0	69,9	74,5	83,5	94,5
7	8,38	12,0	14,1	18,5	24,3	57	62,1	71,0	75,6	84,7	95,8
8	9,52	13,4	15,5	20,1	26,1	58	63,1	72,2	76,8	86,0	97,0
9	10,7	14,7	16,9	21,7	27,9	59	64,2	73,3	77,9	87,2	98,3
10	11,8	16,0	18,3	23,2	29,6	60	65,2	74,4	79,1	88,4	99,6
11	12,9	17,3	19,7	24,7	31,3	61	66,3	75,5	80,2	89,6	100,9
12	14,0	18,5	21,0	26,2	32,9	62	67,3	76,6	81,4	90,8	102,2
13	15,1	19,8	22,4	27,7	34,5	63	68,4	77,7	82,5	92,0	103,4
14	16,2	21,1	23,7	29,1	36,1	64	69,4	78,9	83,7	93,2	104,7
15	17,3	22,3	25,0	30,6	37,7	65	70,5	80,0	84,8	94,4	106,0
16	18,4	23,5	26,3	32,0	39,3	66	71,5	81,1	86,0	95,6	107,3
17	19,5	24,8	27,6	33,4	40,8	67	72,6	82,2	87,1	96,8	108,5
18	20,6	26,0	28,9	34,8	42,3	68	73,6	83,3	88,3	98,0	109,8
19	21,7	27,2	30,1	36,2	43,8	69	74,6	84,4	89,4	99,2	111,1
20	22,8	28,4	31,4	37,6	45,3	70	75,7	85,5	90,5	100,4	112,3
21	23,9	29,6	32,7	38,9	46,8	71	76,7	86,6	91,7	101,6	113,6
22	24,9	30,8	33,9	40,3	48,3	72	77,8	87,7	92,8	102,8	114,8
23	26,0	32,0	35,2	41,6	49,7	73	78,8	88,8	93,9	104,0	116,1
24	27,1	33,2	36,4	43,0	51,2	74	79,9	90,0	95,1	105,2	117,3
25	28,2	34,4	37,7	44,3	52,6	75	80,9	91,1	96,2	106,4	118,6
26	29,2	35,6	38,9	45,6	54,1	76	82,0	92,2	97,4	107,6	119,9
27	30,3	36,7	40,1	47,0	55,5	77	83,0	93,3	98,5	108,8	121,1
28	31,4	37,9	41,3	48,3	56,9	78	84,0	94,4	99,6	110,0	122,3
29	32,5	39,1	42,6	49,6	58,3	79	85,1	95,5	100,7	111,1	123,6
30	33,5	40,3	43,8	50,9	59,7	80	86,1	96,6	101,9	112,3	124,8
31	34,6	41,4	45,0	52,2	61,1	81	87,2	97,7	103,0	113,5	126,1
32	35,7	42,6	46,2	53,5	62,5	82	88,2	98,8	104,1	114,7	127,3
33	36,7	43,7	47,4	54,8	63,9	83	89,2	99,9	105,3	115,9	128,6
34	37,8	44,9	48,6	56,1	65,2	84	90,3	101,0	106,4	117,1	129,8
35	38,9	46,1	49,8	57,3	66,6	85	91,3	102,1	107,5	118,2	131,0
36	39,9	47,2	51,0	58,6	68,0	86	92,4	103,2	108,6	119,4	132,3
37	41,0	48,4	52,2	59,9	69,3	87	93,4	104,3	109,8	120,6	133,5
38	42,0	49,5	53,4	61,2	70,7	88	94,4	105,4	110,9	121,8	134,7
39	43,1	50,7	54,6	62,4	72,1	89	95,5	106,5	112,0	122,9	136,0
40	44,2	51,8	55,8	63,7	73,4	90	96,5	107,6	113,1	124,1	137,2
41	45,2	52,9	56,9	65,0	74,7	91	97,6	108,7	114,3	125,3	138,4
42	46,3	54,1	58,1	66,2	76,1	92	98,6	109,8	115,4	126,5	139,7
43	47,3	55,2	59,3	67,5	77,4	93	99,6	110,9	116,5	127,6	140,9
44	48,4	56,4	60,5	68,7	78,7	94	100,7	111,9	117,6	128,8	142,1
45	49,5	57,5	61,7	70,0	80,1	95	101,7	113,0	118,8	130,0	143,3
46	50,5	58,6	62,8	71,2	81,4	96	102,8	114,1	119,9	131,1	144,6
47	51,6	59,8	64,0	72,4	82,7	97	103,8	115,2	121,0	132,3	145,8
48	52,6	60,9	65,2	73,7	84,0	98	104,8	116,3	122,1	133,5	147,0
49	53,7	62,0	66,3	74,9	85,4	99	105,9	117,4	123,2	134,6	148,2
50	54,7	63,2	67,5	76,2	86,7	100	106,9	118,5	124,3	135,8	149,4

Таблица 11П. Таблица Келли—Вуда (сокращенно)

q	pq	\sqrt{pq}	z		p
0,50	0,2500	0,5000	0,3989	0,0000	0,50
0,49	0,2499	0,4999	0,3988	0,0251	0,51
0,48	0,2496	0,4996	0,3984	0,0502	0,52
0,47	0,2491	0,4991	0,3978	0,0753	0,53
0,46	0,2484	0,4984	0,3969	0,1004	0,54
0,45	0,2475	0,4975	0,3958	0,1257	0,55
0,44	0,2464	0,4964	0,3944	0,1510	0,56
0,43	0,2451	0,4951	0,3928	0,1764	0,57
0,42	0,2436	0,4936	0,3909	0,2019	0,58
0,41	0,2419	0,4918	0,3887	0,2275	0,59
0,40	0,2400	0,4899	0,3863	0,2533	0,60
0,39	0,2379	0,4877	0,3837	0,2793	0,61
0,38	0,2356	0,4854	0,3808	0,3055	0,62
0,37	0,2331	0,4828	0,3776	0,3318	0,63
0,36	0,2304	0,4800	0,3741	0,3585	0,64
0,35	0,2275	0,4770	0,3704	0,3853	0,65
0,34	0,2244	0,4737	0,3664	0,4125	0,66
0,33	0,2211	0,4702	0,3621	0,4399	0,67
0,32	0,2176	0,4665	0,3576	0,4677	0,68
0,31	0,2139	0,4625	0,3529	0,4958	0,69
0,30	0,2100	0,4583	0,3477	0,5244	0,70
0,29	0,2059	0,4538	0,3423	0,5534	0,71
0,28	0,2016	0,4490	0,3366	0,5828	0,72
0,27	0,1971	0,4440	0,3306	0,6128	0,73
0,26	0,1924	0,4386	0,3244	0,6433	0,74
0,25	0,1875	0,4330	0,3178	0,6745	0,75
0,24	0,1824	0,4271	0,3109	0,7063	0,76
0,23	0,1771	0,4208	0,3036	0,7388	0,77
0,22	0,1716	0,4142	0,2961	0,7722	0,78
0,21	0,1659	0,4073	0,2882	0,8064	0,79
0,20	0,1600	0,4000	0,2800	0,8416	0,80
0,19	0,1539	0,3923	0,2714	0,8779	0,81
0,18	0,1476	0,3842	0,2624	0,9154	0,82
0,17	0,1411	0,3756	0,2530	0,9542	0,83
0,16	0,1344	0,3666	0,2433	0,9945	0,84
0,15	0,1275	0,3571	0,2332	1,0364	0,85
0,14	0,1204	0,3470	0,2226	1,0803	0,86
0,13	0,1131	0,3363	0,2115	1,1264	0,87
0,12	0,1056	0,3250	0,2000	1,1750	0,88
0,11	0,0979	0,3129	0,1880	1,2265	0,89
0,10	0,0900	0,3000	0,1755	1,2816	0,90
0,09	0,0819	0,2862	0,1624	1,3408	0,91
0,08	0,0736	0,2713	0,1487	1,4051	0,92
0,07	0,0651	0,2551	0,1343	1,4758	0,93
0,06	0,0564	0,2375	0,1191	1,5548	0,94
0,05	0,0475	0,2179	0,1031	1,6448	0,95
0,04	0,0384	0,1960	0,0862	1,7507	0,96
0,03	0,0291	0,1706	0,0680	1,8808	0,97
0,02	0,0196	0,1400	0,0484	2,0537	0,98
0,01	0,0099	0,0995	0,0266	2,3263	0,99

Таблица 12П. Таблица Сопера для вычисления ошибки бисериального коэффициента корреляции

q	λ_2^2	λ_3	q	λ_2^2	λ_3
0,50	1,5708	2,5000	0,25	1,8567	2,7166
0,49	1,5711	2,5003	0,24	1,8874	2,7399
0,48	1,5722	2,5011	0,23	1,9208	2,7654
0,47	1,5741	2,5024	0,22	1,9574	2,7933
0,46	1,5766	2,5043	0,21	1,9974	2,8240
0,45	1,5799	2,5068	0,20	2,0414	2,8578
0,44	1,5839	2,5098	0,19	2,0898	2,8951
0,43	1,5886	2,5134	0,18	2,1437	2,9364
0,42	1,5943	2,5177	0,17	2,2035	2,9825
0,41	1,6007	2,5225	0,16	2,2703	3,0341
0,40	1,6079	2,5279	0,15	2,3453	3,0923
0,39	1,6161	2,5341	0,14	2,4303	3,1582
0,38	1,6251	2,5409	0,13	2,5272	3,2337
0,37	1,6351	2,5484	0,12	2,6389	3,3208
0,36	1,6461	2,5568	0,11	2,7687	3,4224
0,35	1,6582	2,5659	0,10	2,9221	3,5427
0,34	1,6714	2,5759	0,09	3,1057	3,6873
0,33	1,6858	2,5868	0,08	3,3299	3,8648
0,32	1,7015	2,5986	0,07	3,6110	4,0879
0,31	1,7186	2,6115	0,06	3,9748	4,3777
0,30	1,7371	2,6256	0,05	4,4652	4,7723
0,29	1,7573	2,6409	0,04	5,1715	5,3410
0,28	1,7791	2,6575	0,03	6,2859	6,2485
0,27	1,8028	2,6755	0,02	8,3600	7,9572
0,26	1,8286	2,6952	0,01	13,9393	12,6086

Таблица 13П. Объем выборки, достаточный для достоверности коэффициента корреляции на трех доверительных уровнях *

	N				N			r	N		
	$P'_1 = 0,95$	$P'_2 = 0,99$	$P'_3 = 0,999$		$P'_1 = 0,95$	$P'_2 = 0,99$	$P'_3 = 0,999$		$P'_1 = 0,95$	$P'_2 = 0,99$	$P'_3 = 0,999$
01	38407	66503	108903	31	40	68	109	61	11	16	25
02	9603	16628	27228	32	38	63	102	62	10	16	24
03	4269	7392	12103	33	36	60	96	63	10	15	23
04	2403	4159	6809	34	34	56	90	64	10	15	22
05	1539	2263	4359	35	32	53	85	65	9	14	21
06	1069	1850	3028	36	30	50	80	66	9	14	20
07	787	1360	2225	37	28	47	75	67	9	13	20
08	604	1042	1704	38	27	44	71	68	9	13	19
09	477	824	1347	39	26	42	67	69	8	12	18
10	383	661	1081	40	24	40	64	70	8	12	18
11	317	548	896	41	23	38	60	71	8	11	17
12	267	462	754	42	22	36	57	72	8	11	16
13	228	392	640	43	21	34	55	73	7	11	16
14	196	337	550	44	20	33	52	74	7	10	15
15	171	295	481	45	19	31	49	75	7	10	15
16	151	259	422	46	19	30	47	76	7	10	14
17	133	228	373	47	18	29	45	77	7	9	14
18	119	204	332	48	17	27	43	78	7	9	13
19	107	183	297	49	16	26	41	79	6	9	13
20	97	165	270	50	16	25	39	80	6	9	12
21	87	149	242	51	15	24	37	81	6	8	12
22	80	136	211	52	15	23	36	82	6	8	11
23	73	124	202	53	14	22	34	83	6	8	11
24	68	114	185	54	14	21	33	84	6	7	10
25	62	105	170	55	13	20	32	85	5	7	10
26	57	97	157	56	13	20	30	86	5	7	10
27	53	90	145	57	12	19	29	87	5	7	9
28	49	83	135	58	12	18	28	88	5	7	9
29	46	78	125	59	11	18	27	89	5	6	8
30	43	73	117	60	11	17	26	90	5	6	8

* Коэффициент корреляции приведен в виде первых двух знаков после запятой.

Таблица 14П. Числа Чебышева

5	6	7	8	9	10
$\begin{array}{r} -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ \hline 10 \end{array}$ $\begin{array}{r} +2 \\ -1 \\ -2 \\ -1 \\ +2 \\ \hline 14 \end{array}$ $\begin{array}{r} -1 \\ +2 \\ 0 \\ -2 \\ +1 \\ \hline 10 \end{array}$	$\begin{array}{r} -5 \\ -3 \\ -1 \\ +1 \\ +3 \\ +5 \\ \hline 70 \end{array}$ $\begin{array}{r} +5 \\ -1 \\ -4 \\ -4 \\ -1 \\ +5 \\ \hline 84 \end{array}$ $\begin{array}{r} -5 \\ +7 \\ +4 \\ -4 \\ -7 \\ +5 \\ \hline 180 \end{array}$	$\begin{array}{r} -3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ +3 \\ \hline 28 \end{array}$ $\begin{array}{r} +5 \\ 0 \\ -3 \\ -4 \\ -3 \\ 0 \\ +5 \\ \hline 84 \end{array}$ $\begin{array}{r} -1 \\ +1 \\ +1 \\ 0 \\ -1 \\ -1 \\ +1 \\ \hline 6 \end{array}$	$\begin{array}{r} -7 \\ -5 \\ -3 \\ -1 \\ +1 \\ +3 \\ +5 \\ \hline 168 \end{array}$ $\begin{array}{r} +7 \\ +1 \\ -3 \\ -7 \\ -5 \\ +1 \\ +7 \\ \hline 168 \end{array}$ $\begin{array}{r} -7 \\ +5 \\ +7 \\ -3 \\ -7 \\ -5 \\ +7 \\ \hline 264 \end{array}$	$\begin{array}{r} -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ +3 \\ +4 \\ \hline 60 \end{array}$ $\begin{array}{r} +28 \\ +7 \\ -8 \\ -17 \\ -20 \\ -17 \\ -8 \\ +7 \\ +28 \\ \hline 2772 \end{array}$ $\begin{array}{r} -14 \\ +7 \\ +13 \\ +9 \\ 0 \\ -9 \\ -13 \\ -7 \\ +14 \\ \hline 990 \end{array}$	$\begin{array}{r} -9 \\ -7 \\ -5 \\ -3 \\ -1 \\ +1 \\ +3 \\ +5 \\ +7 \\ +9 \\ \hline 330 \end{array}$ $\begin{array}{r} +6 \\ +2 \\ -1 \\ -3 \\ -4 \\ -4 \\ -3 \\ -1 \\ +2 \\ +6 \\ \hline 132 \end{array}$ $\begin{array}{r} -42 \\ +14 \\ +35 \\ +31 \\ +12 \\ -12 \\ -31 \\ -35 \\ -14 \\ +42 \\ \hline 8580 \end{array}$
11	12	13	14	15	16
$\begin{array}{r} -5 \\ -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ +3 \\ +4 \\ +5 \\ \hline 110 \end{array}$ $\begin{array}{r} +15 \\ +6 \\ -1 \\ -6 \\ -9 \\ -10 \\ -9 \\ -6 \\ -1 \\ +6 \\ +15 \\ \hline 858 \end{array}$ $\begin{array}{r} -30 \\ +6 \\ +22 \\ +23 \\ +14 \\ 0 \\ -14 \\ -23 \\ -22 \\ -6 \\ +30 \\ \hline 4290 \end{array}$	$\begin{array}{r} -11 \\ -9 \\ -7 \\ -5 \\ -3 \\ -1 \\ +1 \\ +3 \\ +5 \\ +7 \\ +9 \\ +11 \\ \hline 572 \end{array}$ $\begin{array}{r} +55 \\ +25 \\ +1 \\ -17 \\ -29 \\ -35 \\ -35 \\ -29 \\ -17 \\ +1 \\ +25 \\ +55 \\ \hline 12012 \end{array}$ $\begin{array}{r} -33 \\ +3 \\ +21 \\ +25 \\ +19 \\ +7 \\ -7 \\ -19 \\ -25 \\ -21 \\ -3 \\ +33 \\ \hline 5148 \end{array}$	$\begin{array}{r} -6 \\ -5 \\ -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ +3 \\ +4 \\ +5 \\ +6 \\ \hline 182 \end{array}$ $\begin{array}{r} +22 \\ +11 \\ +2 \\ -5 \\ -10 \\ -13 \\ -14 \\ -13 \\ -10 \\ -5 \\ -2 \\ +11 \\ +22 \\ \hline 2002 \end{array}$ $\begin{array}{r} -11 \\ 0 \\ +6 \\ +8 \\ +7 \\ +4 \\ 0 \\ -4 \\ -7 \\ -8 \\ -6 \\ 0 \\ +11 \\ \hline 572 \end{array}$	$\begin{array}{r} -13 \\ -11 \\ -9 \\ -7 \\ -5 \\ -3 \\ -1 \\ +1 \\ +3 \\ +5 \\ +7 \\ +9 \\ +11 \\ +13 \\ \hline 910 \end{array}$ $\begin{array}{r} +13 \\ +7 \\ -2 \\ -5 \\ -7 \\ -8 \\ -8 \\ -7 \\ -5 \\ -2 \\ +2 \\ +7 \\ +13 \\ \hline 728 \end{array}$ $\begin{array}{r} -143 \\ -11 \\ +66 \\ +98 \\ +95 \\ +67 \\ +24 \\ -24 \\ -67 \\ -95 \\ -98 \\ -66 \\ +11 \\ +143 \\ \hline 97240 \end{array}$	$\begin{array}{r} -7 \\ -6 \\ -5 \\ -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ +1 \\ +2 \\ +3 \\ +4 \\ +5 \\ +6 \\ +7 \\ \hline 280 \end{array}$ $\begin{array}{r} +91 \\ +52 \\ +19 \\ -8 \\ -29 \\ -44 \\ -53 \\ -56 \\ -53 \\ -44 \\ -29 \\ -8 \\ +19 \\ +52 \\ +91 \\ \hline 37128 \end{array}$ $\begin{array}{r} -91 \\ -13 \\ +35 \\ +58 \\ +61 \\ +49 \\ +27 \\ 0 \\ -27 \\ -49 \\ -61 \\ -58 \\ -35 \\ +13 \\ +91 \\ \hline 39780 \end{array}$	$\begin{array}{r} -15 \\ -13 \\ -11 \\ -9 \\ -7 \\ -5 \\ -3 \\ -1 \\ +1 \\ +3 \\ +5 \\ +7 \\ +9 \\ +11 \\ +13 \\ +15 \\ \hline 1360 \end{array}$ $\begin{array}{r} +35 \\ +21 \\ +9 \\ -1 \\ -9 \\ -15 \\ -19 \\ -21 \\ -21 \\ -19 \\ -15 \\ -9 \\ -1 \\ +9 \\ +21 \\ +35 \\ \hline 5712 \end{array}$ $\begin{array}{r} -455 \\ -91 \\ +143 \\ +267 \\ +301 \\ +265 \\ +179 \\ +63 \\ -63 \\ -179 \\ -265 \\ -301 \\ -267 \\ -143 \\ +91 \\ +455 \\ \hline 100776 \end{array}$

Таблица 14П.

17	18	19	20	21	22
-8	+68	+51	+57	+190	-21
-7	+44	+34	+39	+133	-19
-6	+23	+19	+23	+82	-17
-5	+5	+6	+9	+37	-15
-4	+10	+5	-3	-2	-13
-3	+22	+14	-13	-35	-11
-2	+31	+21	-21	-62	-9
-1	+37	+26	-27	-83	-7
0	+40	+29	-31	-98	-5
+1	+40	+30	-33	-107	-3
+2	+37	+29	-33	-110	-1
+3	+31	+26	-31	-107	+1
+4	+22	+21	-27	-98	+3
+5	+10	+14	-21	-83	+5
+6	+5	+5	-13	-62	+7
+7	+23	+6	-3	-35	+9
+8	+44	+19	+9	-2	+11
408	7752	570	2660	770	3542
3876	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256	243180	4903140	432630	96140
	1938	570	2660	770	3542
	23256	13566	17556	201894	7084
	23256				

23	24	25	26	27
-11	+77	-77	+50	-13
-10	+56	-35	+38	-12
-9	+37	-3	+27	-11
-8	+20	+20	+17	-10
-7	+5	+35	+8	-9
-6	-8	+43	0	-8
-5	-19	+45	-7	-7
-4	-28	+42	-13	-6
-3	-35	-35	-18	-5
-2	-40	+25	-22	-4
-1	-43	+13	-25	-3
0	-44	0	-27	-2
+1	-43	-13	-28	-1
+2	-40	-25	-28	0
+3	-35	-35	-27	+1
+4	-28	-42	-25	+2
+5	-19	-45	-22	+3
+6	-8	-43	-18	+4
+7	+5	-35	-13	+5
+8	+20	-20	-7	+6
+9	+37	+3	0	+7
+10	+56	+35	+8	+8
+11	+77	+77	+17	+9
1012	35420	32890	+27	+10
			+38	+11
			+50	+12
			+5850	+13
			16380	+130
			7803900	712530
				101790

Таблица 14П. (Окончание)

28		29		30		31		32						
-27	+117	-585	-14	+126	-819	-29	+203	-1827	-15	+145	-1015	-31	+155	-899
-25	+91	-325	-13	+99	-463	-27	+161	-1071	-14	+116	-609	-29	+125	-551
-23	+67	-115	-12	+74	-182	-25	+122	-450	-13	+89	-273	-27	+97	-261
-21	+45	+49	-11	+51	+44	-23	+86	+46	-12	+64	-2	-25	+71	-25
-19	+25	+171	-10	+30	+215	-21	+53	+427	-11	+41	+209	-23	+47	+161
-17	+7	+255	-9	+11	+336	-19	+23	+703	-10	+20	+365	-21	+25	+301
-15	-9	+305	-8	-6	+412	-17	-4	+884	-9	+1	+471	-19	+5	+399
-13	-23	+325	-7	-21	+448	-15	-28	+980	-8	-16	+532	-17	-13	+459
-11	-35	+319	-6	-34	+449	-13	-49	+1001	-7	-31	+553	-15	-29	+485
-9	-45	+291	-5	-45	+420	-11	-67	+957	-6	-44	+539	-13	-43	+481
-7	-53	+245	-4	-54	+366	-9	-82	+858	-5	-55	+495	-11	-55	+451
-5	-59	+185	-3	-61	+292	-7	-94	+714	-4	-64	+426	-9	-65	+399
-3	-63	+115	-2	-96	+203	-5	-108	+535	-3	-71	+337	-7	-73	+329
-1	-65	+39	-1	-69	+104	-3	-109	+331	-2	-76	+233	-5	-79	+245
+1	-65	-39	0	-70	0	-1	-112	+112	-1	-79	-119	-3	-93	+151
+3	-63	-115	+1	-69	-104	+1	-112	-112	0	-80	0	-1	-85	+51
+5	-59	-185	+2	-66	-203	+3	-109	-331	+1	-79	-119	+1	-85	-51
+7	-53	-245	+3	-61	-292	+5	-108	-535	+2	-76	-233	+3	-83	-151
+9	-45	-291	+4	-54	-366	+7	-94	-714	+3	-71	-337	+5	-79	-245
+11	-35	-319	+5	-45	-420	+9	-82	-858	+4	-64	-426	+7	-73	-329
+13	-23	-325	+6	-34	-449	+11	-67	-957	+5	-55	-495	+9	-65	-399
+15	-9	-305	+7	-21	-448	+13	-49	-1001	+6	-44	-539	+11	-55	-451
+17	+7	-255	+8	-6	-412	+15	-28	-980	+7	-31	-553	+13	-43	-481
+19	+25	-171	+9	+11	-336	+17	-4	-884	+8	-16	-532	+15	-29	-485
+21	+45	-49	+10	+30	-215	+19	+23	-703	+9	+1	-471	+17	-13	-459
+23	+67	+115	+11	+51	-44	+21	+53	-427	+10	+20	-365	+19	+5	-399
+25	+91	+325	+12	+74	+182	+23	+86	-46	+11	+41	-209	+21	+25	-301
+27	+117	+585	+13	+99	+468	+25	+122	+450	+12	+64	+2	+23	+47	-161
7308	95004	2103660	+14	+126	+819	+27	+161	+1071	+13	+89	+273	+25	+71	+25
			2030	113274	4207320	+29	+203	+1827	+14	+116	+609	+27	+97	+261
			8990	302064	21360240				+15	+145	+1015	+29	+125	+551
									2480	158224	6724520	+31	+155	+899
												10912	185504	5379616

Таблица 15П. Значения $\varphi = 2 \arcsin \sqrt{p}$ (в радианах)

$p\%$	0	1	2	3	4		6	7	8	9
0,0	0,000	0,020	0,028	0,035	0,040	0,045	0,049	0,053	0,057	0,060
0,1	0,063	0,066	0,069	0,072	0,075	0,077	0,080	0,082	0,085	0,087
0,2	0,089	0,092	0,094	0,096	0,098	0,100	0,102	0,104	0,106	0,108
0,3	0,110	0,111	0,113	0,115	0,117	0,118	0,120	0,122	0,123	0,125
0,4	0,127	0,128	0,130	0,131	0,133	0,134	0,136	0,137	0,139	0,140
0,5	0,142	0,143	0,144	0,146	0,147	0,148	0,150	0,151	0,153	0,154
0,6	0,155	0,156	0,158	0,159	0,160	0,161	0,163	0,164	0,165	0,166
0,7	0,168	0,169	0,170	0,171	0,172	0,173	0,175	0,176	0,177	0,178
0,8	0,179	0,180	0,182	0,183	0,184	0,185	0,186	0,187	0,188	0,189
0,9	0,190	0,191	0,192	0,193	0,194	0,195	0,196	0,197	0,198	0,199
1	0,200	0,210	0,220	0,229	0,237	0,246	0,254	0,262	0,269	0,277
2	0,284	0,291	0,298	0,304	0,311	0,318	0,324	0,330	0,336	0,342
3	0,348	0,354	0,360	0,365	0,371	0,376	0,382	0,387	0,392	0,398
4	0,403	0,408	0,413	0,418	0,423	0,428	0,432	0,437	0,442	0,446
5	0,451	0,456	0,460	0,465	0,469	0,473	0,478	0,482	0,486	0,491
6	0,495	0,499	0,503	0,507	0,512	0,516	0,520	0,524	0,528	0,532
7	0,536	0,539	0,543	0,547	0,551	0,555	0,559	0,562	0,566	0,570
8	0,574	0,577	0,581	0,584	0,588	0,592	0,595	0,599	0,602	0,606
9	0,609	0,613	0,616	0,620	0,623	0,627	0,630	0,633	0,637	0,640
10	0,644	0,647	0,650	0,653	0,657	0,660	0,663	0,666	0,670	0,673
11	0,676	0,679	0,682	0,686	0,689	0,692	0,695	0,698	0,701	0,704
12	0,707	0,711	0,714	0,717	0,720	0,723	0,726	0,729	0,732	0,735
13	0,738	0,741	0,744	0,747	0,750	0,752	0,755	0,758	0,761	0,764
14	0,767	0,770	0,773	0,776	0,778	0,781	0,784	0,787	0,790	0,793
15	0,795	0,798	0,801	0,804	0,807	0,809	0,812	0,815	0,818	0,820
16	0,823	0,826	0,828	0,831	0,834	0,837	0,839	0,842	0,845	0,847
17	0,850	0,853	0,855	0,858	0,861	0,863	0,866	0,868	0,871	0,874
18	0,876	0,879	0,881	0,884	0,887	0,889	0,892	0,894	0,897	0,900
19	0,902	0,905	0,907	0,910	0,912	0,915	0,917	0,920	0,922	0,925
20	0,927	0,930	0,932	0,935	0,937	0,940	0,942	0,945	0,947	0,950
21	0,952	0,955	0,957	0,959	0,962	0,964	0,967	0,969	0,972	0,974
22	0,976	0,979	0,981	0,984	0,986	0,988	0,991	0,993	0,996	0,998
23	1,000	1,003	1,005	1,007	1,010	1,012	1,015	1,017	1,019	1,022
24	1,024	1,026	1,029	1,031	1,033	1,036	1,038	1,040	1,043	1,045
25	1,047	1,050	1,052	1,054	1,056	1,059	1,061	1,063	1,066	1,068
26	1,070	1,072	1,075	1,077	1,079	1,082	1,084	1,086	1,088	1,091
27	1,093	1,095	1,097	1,100	1,102	1,104	1,106	1,109	1,111	1,113
28	1,115	1,117	1,120	1,122	1,124	1,126	1,129	1,131	1,133	1,135
29	1,137	1,140	1,142	1,144	1,146	1,148	1,151	1,153	1,155	1,157
30	1,159	1,161	1,164	1,166	1,168	1,170	1,172	1,174	1,177	1,179
31	1,182	1,183	1,185	1,187	1,190	1,192	1,194	1,196	1,198	1,200
32	1,203	1,205	1,207	1,209	1,211	1,213	1,215	1,217	1,220	1,222
33	1,224	1,226	1,228	1,230	1,232	1,234	1,237	1,239	1,241	1,243
34	1,245	1,247	1,249	1,251	1,254	1,256	1,258	1,260	1,262	1,264
35	1,266	1,268	1,270	1,272	1,274	1,277	1,279	1,281	1,283	1,285
36	1,287	1,289	1,291	1,293	1,295	1,297	1,299	1,302	1,304	1,306
37	1,308	1,310	1,312	1,314	1,316	1,318	1,320	1,322	1,324	1,326
38	1,328	1,330	1,333	1,335	1,337	1,339	1,341	1,343	1,345	1,347
39	1,349	1,351	1,353	1,355	1,357	1,359	1,361	1,363	1,365	1,367
40	1,369	1,371	1,374	1,376	1,378	1,380	1,382	1,384	1,386	1,388
41	1,390	1,392	1,394	1,396	1,398	1,400	1,402	1,404	1,406	1,408
42	1,410	1,412	1,414	1,416	1,418	1,420	1,422	1,424	1,426	1,428
43	1,430	1,432	1,434	1,436	1,438	1,440	1,442	1,444	1,446	1,448
44	1,451	1,453	1,455	1,457	1,459	1,461	1,463	1,465	1,467	1,469

Таблица 15П. (Продолжение)

p%	0	1	2	3	4	5	6	7	8	9
45	1,471	1,473	1,475	1,477	1,479	1,481	1,483	1,485	1,487	1,489
46	1,491	1,493	1,495	1,497	1,499	1,501	1,503	1,505	1,507	1,509
47	1,511	1,513	1,515	1,517	1,519	1,521	1,523	1,525	1,527	1,529
48	1,531	1,533	1,535	1,537	1,539	1,541	1,543	1,545	1,547	1,549
49	1,551	1,553	1,555	1,557	1,559	1,561	1,563	1,565	1,567	1,569
50	1,571	1,573	1,575	1,577	1,579	1,581	1,583	1,585	1,587	1,589
51	1,591	1,593	1,595	1,597	1,599	1,601	1,603	1,605	1,607	1,609
52	1,611	1,613	1,615	1,617	1,619	1,621	1,623	1,625	1,627	1,629
53	1,631	1,633	1,635	1,637	1,639	1,641	1,643	1,645	1,647	1,649
54	1,651	1,653	1,655	1,657	1,659	1,661	1,663	1,665	1,667	1,669
55	1,671	1,673	1,675	1,677	1,679	1,681	1,683	1,685	1,687	1,689
56	1,691	1,693	1,695	1,697	1,699	1,701	1,703	1,705	1,707	1,709
57	1,711	1,713	1,715	1,717	1,719	1,721	1,723	1,725	1,727	1,729
58	1,731	1,734	1,736	1,738	1,740	1,742	1,744	1,746	1,748	1,750
59	1,752	1,754	1,756	1,758	1,760	1,762	1,764	1,766	1,768	1,770
60	1,772	1,774	1,776	1,778	1,780	1,782	1,784	1,786	1,789	1,791
61	1,793	1,795	1,797	1,799	1,801	1,803	1,805	1,807	1,809	1,811
62	1,813	1,815	1,817	1,819	1,821	1,823	1,826	1,828	1,830	1,832
63	1,834	1,836	1,838	1,840	1,842	1,844	1,846	1,848	1,850	1,853
64	1,855	1,857	1,857	1,861	1,863	1,865	1,867	1,869	1,871	1,873
65	1,875	1,878	1,880	1,882	1,884	1,886	1,888	1,890	1,892	1,894
66	1,897	1,899	1,901	1,903	1,905	1,907	1,909	1,911	1,913	1,916
67	1,918	1,920	1,922	1,924	1,926	1,928	1,930	1,933	1,935	1,937
68	1,939	1,941	1,943	1,946	1,948	1,950	1,952	1,954	1,956	1,958
69	1,961	1,963	1,865	1,967	1,969	1,971	1,974	1,976	1,978	1,980
70	1,982	1,984	1,987	1,989	1,991	1,993	1,995	1,998	2,000	2,002
71	2,004	2,006	2,009	2,011	2,013	2,015	2,018	2,020	2,022	2,024
72	2,026	2,029	2,031	2,033	2,035	2,038	2,040	2,042	2,044	2,047
73	2,049	2,051	2,053	2,056	2,058	2,060	2,062	2,065	2,067	2,069
74	2,071	2,074	2,076	2,078	2,081	2,083	2,085	2,087	2,090	2,092
75	2,094	2,097	2,099	2,101	2,104	2,106	2,108	2,111	2,113	2,115
76	2,118	2,120	2,122	2,125	2,127	2,129	2,132	2,134	2,136	2,139
77	2,141	2,144	2,146	2,148	2,151	2,153	2,156	2,158	2,160	2,163
78	2,165	2,168	2,170	2,172	2,175	2,177	2,180	2,182	2,185	2,187
79	2,190	2,192	2,194	2,197	2,199	2,202	2,204	2,207	2,209	2,212
80	2,214	2,217	2,219	2,222	2,224	2,227	2,229	2,231	2,234	2,237
81	2,240	2,242	2,245	2,247	2,250	2,252	2,255	2,258	2,260	2,263
82	2,265	2,268	2,271	2,273	2,276	2,278	2,281	2,284	2,286	2,289
83	2,292	2,294	2,297	2,300	2,302	2,305	2,308	2,310	2,313	2,316
84	2,319	2,321	2,324	2,327	2,330	2,332	2,335	2,338	2,341	2,343
85	2,346	2,349	2,352	2,355	2,357	2,360	2,363	2,366	2,369	2,372
86	2,375	2,377	2,380	2,383	2,386	2,389	2,392	2,395	2,398	2,401

Таблица 15П. (Окончание)

p%	0	1	2	3	4	5		7	8	9
87	2,404	2,407	2,410	2,413	2,416	2,419	2,422	2,425	2,428	2,431
88	2,434	2,437	2,440	2,443	2,447	2,450	2,453	2,456	2,459	2,462
89	2,465	2,469	2,472	2,475	2,478	2,482	2,485	2,488	2,491	2,495
90	2,498	2,501	2,505	2,508	2,512	2,515	2,518	2,522	2,525	2,529
91	2,532	2,536	2,539	2,543	2,546	2,550	2,554	2,557	2,561	2,564
92	2,568	2,572	2,575	2,579	2,583	2,587	2,591	2,594	2,598	2,602
93	2,606	2,610	2,614	2,618	2,622	2,626	2,630	2,634	2,638	2,642
94	2,647	2,651	2,655	2,659	2,664	2,668	2,673	2,677	2,681	2,686
95	2,691	2,695	2,700	2,705	2,709	2,714	2,719	2,724	2,729	2,734
96	2,739	2,744	2,749	2,754	2,760	2,765	2,771	2,776	2,782	2,788
97	2,793	2,799	2,805	2,811	2,818	2,824	2,830	2,837	2,844	2,851
98	2,858	2,865	2,872	2,880	2,888	2,896	2,904	2,913	2,922	2,931
99,0	2,941	2,942	2,943	2,944	2,945	2,946	2,948	2,949	2,950	2,951
99,1	2,952	2,953	2,954	2,955	2,956	2,957	2,958	2,959	2,960	2,961
99,2	2,963	2,964	2,965	2,966	2,967	2,968	2,969	2,971	2,972	2,973
99,3	2,974	2,975	2,976	2,978	2,979	2,980	2,981	2,983	2,984	2,985
99,4	2,987	2,988	2,989	2,990	2,992	2,993	2,995	2,996	2,997	2,999
99,5	3,000	3,002	3,003	3,004	3,006	3,007	3,009	3,010	3,012	3,013
99,6	3,015	3,017	3,018	3,020	3,022	3,023	3,025	3,027	3,028	3,030
99,7	3,032	3,034	3,036	3,038	3,040	3,041	3,044	3,046	3,048	3,050
99,8	3,052	3,054	3,057	3,059	3,062	3,064	3,067	3,069	3,072	3,075
99,9	3,078	3,082	3,085	3,089	3,093	3,097	3,101	3,107	3,113	3,122
100	3,142									

Таблица 16П. Граничные значения критерия Вилкоксона *

N_y	N_x	4	5	6	7	8	9	10					
4	10	11	—	12	10	13	10	14	11	15	11	15	12
5		17	15	18	16	20	17	21	17	22	18	23	19
6				26	23	27	24	29	25	31	26	32	27
7						36	32	38	34	40	35	42	37
8								49	43	51	45	53	47
9										63	56	65	58
10												78	71

* Нулевая гипотеза принимается при n больше левого числа (при $W_1' = 0,05$) и отвергается при n меньше правого числа в той же колонке (при $W_2' = 0,01$).

Таблица 17П. Значения $\Psi\left(\frac{x_i}{N+1}\right)$ -функции, обратной к интегралу вероятностей

$\frac{x_i}{N+1}$	0	1	2	3	4	5	6	7	8	9
0,00	$-\infty$	-3,09	-2,88	-2,75	-2,65	-2,58	-2,51	-2,46	-2,41	-2,37
0,01	-2,33	-2,29	-2,26	-2,23	-2,20	-2,17	-2,14	-2,12	-2,10	-2,07
0,02	-2,05	-2,03	-2,01	-2,00	-1,98	-1,96	-1,94	-1,93	-1,91	-1,90
0,03	-1,88	-1,87	-1,85	-1,84	-1,83	-1,81	-1,80	-1,79	-1,77	-1,76
0,04	-1,75	-1,74	-1,73	-1,72	-1,71	-1,70	-1,68	-1,67	-1,66	-1,65
0,05	-1,64	-1,64	-1,63	-1,62	-1,61	-1,60	-1,59	-1,58	-1,57	-1,56
0,06	-1,55	-1,55	-1,54	-1,53	-1,52	-1,51	-1,51	-1,50	-1,49	-1,48
0,07	-1,48	-1,47	-1,46	-1,45	-1,45	-1,44	-1,43	-1,43	-1,42	-1,41
0,08	-1,41	-1,40	-1,39	-1,39	-1,38	-1,37	-1,37	-1,36	-1,35	-1,35
0,09	-1,34	-1,33	-1,33	-1,32	-1,32	-1,31	-1,30	-1,30	-1,29	-1,29
0,10	-1,28	-1,28	-1,27	-1,26	-1,26	-1,25	-1,25	-1,24	-1,24	-1,23
0,11	-1,28	-1,22	-1,22	-1,21	-1,21	-1,20	-1,20	-1,19	-1,19	-1,18
0,12	-1,18	-1,17	-1,17	-1,16	-1,16	-1,15	-1,15	-1,14	-1,14	-1,13
0,13	-1,13	-1,12	-1,12	-1,11	-1,11	-1,10	-1,10	-1,09	-1,09	-1,09
0,14	-1,08	-1,08	-1,07	-1,07	-1,06	-1,06	-1,05	-1,05	-1,05	-1,04
0,15	-1,04	-1,03	-0,03	-1,02	-1,02	-1,02	-1,01	-1,01	-1,00	-1,00
0,16	-0,99	-0,99	-0,99	-0,98	-0,98	-0,97	-0,97	-0,97	-0,96	-0,96
0,17	-0,95	-0,95	-0,95	-0,94	-0,94	-0,93	-0,93	-0,93	-0,92	-0,92
0,18	-0,92	-0,91	-0,91	-0,90	-0,90	-0,90	-0,89	-0,89	-0,89	-0,88
0,19	-0,88	-0,87	-0,87	-0,87	-0,86	-0,86	-0,86	-0,85	-0,85	-0,85
0,20	-0,84	-0,84	-0,83	-0,83	-0,83	-0,82	-0,82	-0,82	-0,81	-0,81
0,21	-0,81	-0,80	-0,80	-0,80	-0,79	-0,79	-0,79	-0,78	-0,78	-0,78
0,22	-0,77	-0,77	-0,77	-0,76	-0,76	-0,76	-0,75	-0,75	-0,75	-0,74
0,23	-0,74	-0,74	-0,73	-0,73	-0,73	-0,72	-0,72	-0,72	-0,71	-0,71
0,24	-0,71	-0,70	-0,70	-0,70	-0,69	-0,69	-0,69	-0,68	-0,68	-0,68
0,25	-0,67	-0,67	-0,67	-0,67	-0,66	-0,66	-0,66	-0,65	-0,65	-0,65
0,26	-0,64	-0,64	-0,64	-0,63	-0,63	-0,63	-0,63	-0,62	-0,62	-0,62
0,27	-0,61	-0,61	-0,61	-0,60	-0,60	-0,60	-0,59	-0,59	-0,59	-0,59
0,28	-0,58	-0,58	-0,58	-0,57	-0,57	-0,57	-0,57	-0,56	-0,56	-0,56
0,29	-0,55	-0,55	-0,55	-0,54	-0,54	-0,54	-0,54	-0,53	-0,53	-0,53
0,30	-0,52	-0,52	-0,52	-0,52	-0,51	-0,51	-0,51	-0,50	-0,50	-0,50
0,31	-0,50	-0,49	-0,49	-0,49	-0,48	-0,48	-0,48	-0,48	-0,47	-0,47
0,32	-0,47	-0,46	-0,46	-0,46	-0,46	-0,45	-0,45	-0,45	-0,45	-0,44
0,33	-0,44	-0,44	-0,43	-0,43	-0,43	-0,43	-0,42	-0,42	-0,42	-0,42
0,34	-0,41	-0,41	-0,41	-0,40	-0,40	-0,40	-0,40	-0,39	-0,39	-0,39
0,35	-0,39	-0,38	-0,38	-0,38	-0,37	-0,37	-0,37	-0,37	-0,36	-0,36
0,36	-0,36	-0,36	-0,35	-0,35	-0,35	-0,35	-0,34	-0,34	-0,34	-0,33
0,37	-0,33	-0,33	-0,33	-0,32	-0,32	-0,32	-0,32	-0,31	-0,31	-0,31
0,38	-0,31	-0,30	-0,30	-0,30	-0,30	-0,29	-0,29	-0,29	-0,28	-0,28
0,39	-0,28	-0,28	-0,27	-0,27	-0,27	-0,27	-0,26	-0,26	-0,26	-0,26
0,40	-0,25	-0,25	-0,25	-0,25	-0,24	-0,24	-0,24	-0,24	-0,23	-0,23
0,41	-0,23	-0,23	-0,22	-0,22	-0,22	-0,21	-0,21	-0,21	-0,21	-0,20
0,42	-0,20	-0,20	-0,20	-0,19	-0,19	-0,19	-0,19	-0,18	-0,18	-0,18
0,43	-0,18	-0,17	-0,17	-0,17	-0,17	-0,16	-0,16	-0,16	-0,16	-0,15
0,44	-0,15	-0,15	-0,15	-0,14	-0,14	-0,14	-0,14	-0,13	-0,13	-0,13
0,45	-0,13	-0,12	-0,12	-0,12	-0,12	-0,11	-0,11	-0,11	-0,11	-0,10
0,46	-0,10	-0,10	-0,10	-0,09	-0,09	-0,09	-0,09	-0,08	-0,08	-0,08
0,47	-0,08	-0,07	-0,07	-0,07	-0,07	-0,06	-0,06	-0,06	-0,06	-0,05
0,48	-0,05	-0,05	-0,05	-0,04	-0,04	-0,04	-0,04	-0,03	-0,03	-0,03

Таблица 17П. (Окончание)

$\frac{x_i}{N+1}$	0	1	2	3	4	5	6	7	8	9
0,49	-0,03	-0,02	-0,02	-0,02	-0,02	-0,01	-0,01	-0,01	-0,01	-0,00
0,50	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,02
0,51	0,03	0,03	0,03	0,03	0,04	0,04	0,04	0,04	0,05	0,05
0,52	0,05	0,05	0,06	0,06	0,06	0,06	0,07	0,07	0,07	0,07
0,53	0,08	0,08	0,08	0,08	0,09	0,09	0,09	0,09	0,10	0,10
0,54	0,10	0,10	0,11	0,11	0,11	0,11	0,12	0,12	0,12	0,12
0,55	0,13	0,13	0,13	0,13	0,14	0,14	0,14	0,14	0,15	0,15
0,56	0,15	0,15	0,16	0,16	0,16	0,16	0,17	0,17	0,17	0,17
0,57	0,18	0,18	0,18	0,18	0,19	0,19	0,19	0,19	0,20	0,20
0,58	0,20	0,20	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23
0,59	0,23	0,23	0,23	0,24	0,24	0,24	0,24	0,25	0,25	0,25
0,60	0,25	0,26	0,26	0,26	0,26	0,27	0,27	0,27	0,27	0,28
0,61	0,28	0,28	0,28	0,29	0,29	0,29	0,30	0,30	0,30	0,30
0,62	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32	0,33	0,33
0,63	0,33	0,33	0,34	0,34	0,34	0,35	0,35	0,35	0,35	0,36
0,64	0,36	0,36	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38
0,65	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41
0,66	0,41	0,42	0,42	0,42	0,42	0,43	0,43	0,43	0,43	0,44
0,67	0,44	0,44	0,45	0,45	0,45	0,45	0,46	0,46	0,46	0,46
0,68	0,47	0,47	0,47	0,48	0,48	0,48	0,48	0,49	0,49	0,49
0,69	0,50	0,50	0,50	0,50	0,51	0,51	0,51	0,52	0,52	0,52
0,70	0,52	0,53	0,53	0,53	0,54	0,54	0,54	0,54	0,55	0,55
0,71	0,55	0,56	0,56	0,56	0,57	0,57	0,57	0,57	0,58	0,58
0,72	0,58	0,59	0,59	0,59	0,59	0,60	0,60	0,60	0,61	0,61
0,73	0,61	0,62	0,62	0,62	0,63	0,63	0,63	0,63	0,64	0,64
0,74	0,64	0,65	0,65	0,65	0,66	0,66	0,66	0,67	0,67	0,67
0,75	0,67	0,68	0,68	0,68	0,69	0,69	0,69	0,70	0,70	0,70
0,76	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,74
0,77	0,74	0,74	0,75	0,75	0,75	0,76	0,76	0,76	0,77	0,77
0,78	0,77	0,78	0,78	0,78	0,79	0,79	0,79	0,80	0,80	0,80
0,79	0,81	0,81	0,81	0,82	0,82	0,82	0,83	0,83	0,83	0,84
0,80	0,84	0,85	0,85	0,85	0,86	0,86	0,86	0,87	0,87	0,87
0,81	0,88	0,88	0,89	0,89	0,89	0,90	0,90	0,90	0,91	0,91
0,82	0,92	0,92	0,92	0,93	0,93	0,93	0,94	0,94	0,95	0,95
0,83	0,95	0,96	0,96	0,97	0,97	0,97	0,98	0,98	0,99	0,99
0,84	0,99	1,00	1,00	1,01	1,01	1,02	1,02	1,02	1,03	1,03
0,85	1,04	1,04	1,05	1,05	1,05	1,06	1,06	1,07	1,07	1,08
0,86	1,08	1,09	1,09	1,09	1,10	1,10	1,11	1,11	1,12	1,12
0,87	1,13	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,17	1,17
0,88	1,18	1,18	1,19	1,19	1,20	1,20	1,21	1,21	1,22	1,22
0,89	1,23	1,23	1,24	1,24	1,25	1,25	1,26	1,26	1,27	1,27
0,90	1,28	1,29	1,29	1,30	1,30	1,31	1,32	1,32	1,33	1,33
0,91	1,34	1,35	1,35	1,36	1,37	1,37	1,38	1,39	1,39	1,40
0,92	1,41	1,41	1,42	1,43	1,43	1,44	1,45	1,45	1,46	1,47
0,93	1,48	1,48	1,49	1,50	1,51	1,51	1,52	1,53	1,54	1,55
0,94	1,55	1,56	1,57	1,58	1,59	1,60	1,61	1,62	1,63	1,64
0,95	1,64	1,65	1,66	1,67	1,68	1,70	1,71	1,72	1,73	1,74
0,96	1,75	1,76	1,77	1,79	1,80	1,81	1,83	1,84	1,85	1,87
0,97	1,88	1,90	1,91	1,93	1,94	1,96	1,98	2,00	2,01	2,03
0,98	2,05	2,07	2,10	2,12	2,14	2,17	2,20	2,23	2,26	2,29
0,99	2,33	2,37	2,41	2,46	2,51	2,58	2,65	2,75	2,88	3,09

Таблица 18П. Граничные значения критерия ван дер Вардена

N	$N_x - N_y = 0$ или 1		$N_x - N_y = 2$ или 3		$N_x - N_y = 4$ или 5		N	$N_x - N_y = 0$ или 1		$N_x - N_y = 2$ или 3		$N_x - N_y = 4$ или 5	
	$W_1 = 5\%$	$W_2 = 1\%$	$W_1 = 5\%$	$W_2 = 1\%$	$W_1 = 5\%$	$W_2 = 1\%$		$W_1 = 5\%$	$W_2 = 1\%$	$W_1 = 5\%$	$W_2 = 1\%$	$W_1 = 5\%$	$W_2 = 1\%$
6	∞	∞	∞	∞	∞	∞	29	4,78	6,22	4,76	6,19	4,72	6,13
7	∞	∞	∞	∞	∞	∞	30	4,88	6,35	4,87	6,34	4,84	6,30
8	2,40	∞	2,30	∞	∞	∞	31	4,97	6,47	4,95	6,44	4,91	6,39
9	2,38	∞	2,20	∞	∞	∞	32	5,07	6,60	5,06	6,58	5,03	6,55
10	2,60	3,20	2,49	3,10	2,30	∞	33	5,15	6,71	5,13	6,69	5,10	6,64
11	2,72	3,40	2,58	3,40	2,40	∞	34	5,25	6,84	5,24	6,82	5,21	6,79
12	2,86	3,60	2,79	3,58	2,68	3,40	35	5,33	6,95	5,31	6,92	5,28	6,88
13	2,96	3,71	2,91	3,64	2,78	3,50	36	5,42	7,06	5,41	7,05	5,38	7,02
14	3,11	3,94	3,06	3,88	3,00	3,76	37	5,50	7,17	5,48	7,15	5,45	7,11
15	3,24	4,07	3,19	4,05	3,06	3,88	38	5,59	7,28	5,58	7,27	5,55	7,25
16	3,39	4,26	3,36	4,25	3,28	4,12	39	5,67	7,39	5,65	7,37	5,62	7,33
17	3,49	4,44	3,44	4,37	3,36	4,23	40	5,75	7,50	5,74	7,49	5,72	7,47
18	3,63	4,60	3,60	4,58	3,53	4,50	41	5,83	7,62	5,81	7,60	5,79	7,56
19	3,73	4,77	3,69	4,71	3,61	4,62	42	5,91	7,72	5,90	7,71	5,88	7,69
20	3,86	4,94	3,84	4,92	3,78	4,85	43	5,99	7,82	5,97	7,81	5,95	7,77
21	3,96	5,10	3,92	5,05	3,85	4,96	44	6,06	7,93	6,06	7,92	6,04	7,90
22	4,08	5,26	4,06	5,24	4,01	5,17	45	6,14	8,02	6,12	8,01	6,10	7,98
23	4,18	5,40	4,15	5,36	4,08	5,27	46	6,21	8,13	6,21	8,12	6,19	8,10
24	4,29	5,55	4,27	5,53	4,23	5,48	47	6,29	8,22	6,27	8,21	6,25	8,18
25	4,39	5,68	4,36	5,65	4,30	5,58	48	6,36	8,32	6,35	8,31	6,34	8,29
26	4,50	5,83	4,48	5,81	4,44	5,76	49	6,43	8,41	6,42	8,40	6,39	8,37
27	4,59	5,95	4,56	5,92	4,51	5,85	50	6,50	8,51	6,50	8,50	6,48	8,48
28	4,69	6,09	4,68	6,07	4,64	6,03							

Таблица 19П. Граничные значения числа серий

N_x	N_y	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6		3	3	3														
7		3	3	4	4													
8		3	3	4	4	5												
9		3	4	4	5	5	6											
10		3	4	5	5	6	6	6										
11		3	4	5	5	6	6	7	7									
12		4	4	5	6	6	7	7	8	8								
13		4	4	5	6	6	7	8	8	9	9							
14		4	5	5	6	7	7	8	8	9	9	10						
15		4	5	6	6	7	8	8	9	9	10	10	11					
16		4	5	6	6	7	8	8	9	10	10	11	11	11				
17		4	5	6	7	7	8	9	9	10	10	11	11	12	12			
18		4	5	6	7	8	8	9	10	10	11	11	12	12	13	13		
19		4	5	6	7	8	8	9	10	10	11	12	12	13	13	14	14	
20		4	5	6	7	8	9	9	10	11	11	12	12	13	13	14	14	15

Таблица 20П. Граничные значения числа знаков (менее часто встречающихся) при двух уровнях значимости

<i>N</i>	5%	1%	<i>N</i>	5%	1%	<i>N</i>	5%	1%	<i>N</i>	5%	1%
8	1	0	31	10	7	54	20	17	77	30	26
9	2	0	32	10	8	55	20	17	78	30	27
10	2	0	33	11	8	56	21	17	79	31	27
11	2	0	34	11	9	57	21	18	80	31	28
12	3	1	35	12	9	58	22	18	81	32	28
13	3	1	36	12	9	59	22	19	82	32	28
14	3	1	37	13	10	60	22	19	83	33	29
15	4	2	38	13	10	61	23	20	84	33	29
16	4	2	39	13	11	62	23	20	85	33	30
17	5	2	40	14	11	63	24	20	86	34	30
18	5	3	41	14	11	64	24	21	87	34	31
19	5	3	42	15	12	65	25	21	88	35	31
20	6	3	43	15	12	66	25	22	89	35	31
21	6	4	44	16	13	67	26	22	90	36	32
22	6	4	45	16	13	68	26	22	91	36	32
23	7	4	46	16	13	69	26	23	92	37	33
24	7	5	47	17	14	70	27	23	93	37	33
25	8	5	48	17	14	71	27	24	94	38	34
26	8	6	49	18	15	72	28	24	95	38	34
27	8	6	50	18	15	73	28	25	96	38	34
28	9	6	51	19	15	74	29	25	97	39	35
29	9	7	52	19	16	75	29	25	98	39	35
30	10	7	53	19	16	76	29	26	99	40	36
									100	40	36

Таблица 21П. Граничные значения критерия Вилкоксона для рядов с попарно связанными вариантами *

<i>N</i>	$W_1 = 5\%$	$W_2 = 1\%$	<i>N</i>	$W_1 = 5\%$	$W_2 = 1\%$	<i>N</i>	$W_1 = 5\%$	$W_2 = 1\%$
6	0	—	13	17	10	20	52	38
7	2	—	14	21	13	21	59	43
8	4	0	15	25	16	22	66	49
9	6	2	16	30	20	23	73	55
10	8	3	17	35	23	24	81	61
11	11	5	18	40	28	25	89	68
12	14	7	19	46	32			

* Нулевая гипотеза принимается, если вычисленная сумма рангов (n_T) будет больше табличной при $W_1 = 5\%$ и отвергается, если сумма рангов будет меньше табличной при $W_2 = 1\%$.

Таблица 22П. Показательная функция e^{-x} (нуль целых опущен)

	0	1	2	3	4	5	6	7	8	9
0,00	1,	9990	9980	9970	9960	9950	9940	9930	9920	9910
0,0	1,	9900	9802	9704	9608	9512	9418	9324	9231	9139
1	9048	8958	8869	8781	8694	8607	8521	8437	8353	8270
2	8187	8106	8025	7945	7866	7788	7711	7634	7558	7483
3	7408	7334	7261	7189	7118	7047	6977	6907	6839	6771
4	6703	6637	6570	6505	6440	6376	6313	6250	6188	6126
0,5	6065	6005	5945	5886	5827	5769	5712	5655	5599	5543
6	5488	5434	5379	5326	5273	5220	5169	5117	5066	5016
7	4966	4916	4868	4819	4771	4724	4677	4630	4584	4538
8	4493	4449	4404	4360	4317	4274	4232	4190	4148	4107
9	4066	4025	3985	3946	3906	3867	3829	3791	3753	3716
1,	3679	3329	3012	2725	2466	2231	2019	1827	1653	1496
2,	1353	1225	1108	1003	0907	0821	0743	0672	0608	0550
3,	0498	0450	0408	0369	0334	0302	0273	0247	0224	0202
4,	0183	0166	0150	0136	0123	0111	0100	0091	0082	0074
5,	0067	0061	0055	0050	0045	0041	0037	0034	0030	0027
6,	0025	0022	0020	0018	0017	0015	0014	0012	0011	0010
7,	0009	0008	0007	0007	0006	0006	0005	0005	0004	0004
8,	0003	0003	0003	0002	0002	0002	0002	0002	0002	0001
9,	0001	0001	0001	0001	0001	0001	0001	0001	0001	0000

Таблица 23П. Минимальное число наблюдений в зависимости от варьирования данных, при доверительном уровне $P_1 = 95\%$ и показателе точности 5%

Коэффициент вариации, v, %	0	1	2	3	4	5	6	7	8	9
0	5	5	5	5	5	5	6	8	10	12
10	15	19	22	26	30	35	39	44	50	55
20	61	68	74	81	89	96	104	112	121	129
30	138	148	157	167	178	188	199	210	222	234
40	246	258	271	284	298	311	325	340	354	369
50	384	400	416	432	448	465	482	499	517	535
60	553	572	591	610	630	649	670	690	711	732

Таблица 24П. Перевод календарных дат в непрерывный ряд

Месяцы											
III	IV	V	VI	VII	VIII	IX	X	XI	XII	I	II
1	32	62	93	123	154	185	215	246	276	307	338
2	33	63	94	124	155	186	216	247	277	308	339
3	34	64	95	125	156	187	217	248	278	309	340
4	35	65	96	126	157	188	218	249	279	310	341
5	36	66	97	127	158	189	219	250	280	311	342
6	37	67	98	128	159	190	220	251	281	312	343
7	38	68	99	129	160	191	221	252	282	313	344
8	39	69	100	130	161	192	222	253	283	314	345
9	40	70	101	131	162	193	223	254	284	315	346
10	41	71	102	132	163	194	224	255	285	316	347
11	42	72	103	133	164	195	225	256	286	317	348
12	43	73	104	134	165	196	226	257	287	318	349
13	44	74	105	135	166	197	227	258	288	319	350
14	45	75	106	136	167	198	228	259	289	320	351
15	46	76	107	137	168	199	229	260	290	321	352
16	47	77	108	138	169	200	230	261	291	322	353
17	48	78	109	139	170	201	231	262	292	323	354
18	49	79	110	140	171	202	232	263	293	324	355
19	50	80	111	141	172	203	233	264	294	25	356
20	51	81	112	142	173	204	234	265	295	326	357
21	52	82	113	143	174	205	235	266	296	327	358
22	53	83	114	144	175	206	236	267	297	328	359
23	54	84	115	145	176	207	237	268	298	329	360
24	55	85	116	146	177	208	238	269	299	330	361
25	56	86	117	147	178	209	239	270	300	331	362
26	57	87	118	148	179	210	240	271	301	332	363
27	58	88	119	149	180	211	241	272	302	333	364
28	59	89	120	150	181	212	242	273	303	334	365
29	60	90	121	151	182	213	243	274	304	335	(366)
30	61	91	122	152	183	214	244	275	305	336	—
31	—	92	—	153	184	—	245	—	306	337	—

Таблица 25П. Значения критерия Стьюдента для односторонней оценки гипотез

Число степеней свободы, ν	Доверительные уровни			Число степеней свободы, ν	Доверительные уровни			Число степеней свободы, ν	Доверительные уровни		
	95%	99%	99,9%		95%	99%	99,9%		95%	99%	99,9%
1	6,31	31,82	318,31	13	1,77	2,65	3,85	24	1,71	2,49	3,47
2	2,92	6,97	22,33	14	1,76	2,62	3,79	25	1,71	2,49	3,45
3	2,35	4,54	10,21	15	1,75	2,60	3,73	26	1,71	2,48	3,44
4	2,13	3,75	7,17	16	1,75	2,58	3,69	27	1,70	2,47	3,42
5	2,01	3,37	5,89	17	1,74	2,57	3,65	28	1,70	2,47	3,41
6	1,94	3,14	5,21	18	1,73	2,55	3,61	29	1,70	2,46	3,40
7	1,89	3,00	4,79	19	1,73	2,54	3,58	30	1,70	2,46	3,39
8	1,86	2,90	4,50	20	1,73	2,53	3,55	40	1,68	2,42	3,31
9	1,83	2,82	4,30	21	1,72	2,52	3,53	60	1,67	2,39	3,23
10	1,81	2,76	4,14	22	1,72	2,51	3,51	120	1,66	2,36	3,17
11	1,80	2,72	4,03	23	1,71	2,50	3,49	∞	1,64	2,33	3,09
12	1,78	2,68	3,93								

Таблица 26П. Функция $z = 0,5 \ln \frac{1+r}{1-r}$

r	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,1000	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Примечание. Обратное преобразование делается с помощью таблиц гиперболического тгенса ($r = \text{th } z$) или по формуле $r = \frac{e^{2z} - 1}{e^{2z} + 1}$.

Достоверность коэффициента корреляции (r) можно оценить, не обращаясь к таблице 26П по формуле

$$t = \left(0,5 \ln \frac{1+r}{1-r} \right) \sqrt{N-3}, \quad (5.43)$$

где N — объем ряда, t — критерий Стьюдента, при числе степеней свободы $\nu = N - 2$ (табл. 3П).

Таблица 27П. Функция $\Phi(x)$ нормированного нормального распределения (перед числами в таблице поставить: 0,)

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54379	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173

Таблица 27П. (Окончание)

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0,5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84849	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89616	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2,4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
2,5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2,9	99813	99819	99825	99830	99836	99841	99846	99851	99856	99860
x	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3,0	99865	99903	99931	99952	99966	99977	99984	99989	99993	99995

Таблица 28П. Значения $1 - K(\lambda)$ критерия Колмогорова

λ	0	1	2	3	4	5	6	7	8	9
0,3	99999	99998	99995	99991	99983	99970	99949	99917	99872	99807
0,4	99719	99603	99452	99262	99027	98741	98400	97998	97532	96998
0,5	96394	95719	94969	94147	93250	92282	91242	90134	88960	87724
0,6	86428	85077	83678	82225	80732	79201	77636	76042	74422	72781
0,7	71124	69453	67774	66089	64402	62717	61036	59363	57700	56050
0,8	54414	52796	51197	49619	48063	46532	45026	43545	42093	40668
0,9	39273	37907	36571	35266	33992	32748	31536	30356	29206	28087
1,0	27000	25943	24917	23922	22957	22021	21114	20236	19387	18566
1,1	17772	17005	16264	15550	14861	14196	13556	12939	12345	11774
1,2	11225	10697	10190	09703	09235	08787	08357	07944	07550	07171
1,3	06809	06463	06132	05815	05513	05224	04949	04686	04435	04196
1,4	03968	03751	03545	03348	03162	02984	02815	02655	02503	02359
1,5	02222	02092	01969	01852	01742	01638	01539	01446	01357	01274
1,6	01195	01121	01051	00985	00922	00864	00808	00756	00707	00661
1,7	00618	00577	00539	00503	00469	00438	00408	00380	00354	00330
1,8	00307	00285	00265	00247	00229	00213	00198	00186	00170	00158
1,9	00146	00136	00126	00116	00108	00100	00092	00085	00079	00073
2,0	00067	00062	00057	00053	00048	00045	00041	00038	00035	00032
2,1	00030	00027	00025	00023	00021	00019	00018	00016	00015	00014
2,2	00013	00011	00010	00010	00009	00008	00007	00007	00006	00006
2,3	00005	00005	00004	00004	00004	00003	00003	00003	00002	00002
2,4	00002	00002	00002	00001	00001	00001	00001	00001	00001	00001

Таблица 29П. Множители для вычисления коэффициентов уравнений парабол (нуль целых опущен, в скобках число нулей между запятой и первой цифрой мантиссы)

N	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈
5	2	1	4857143	1428571	0714286	9027778	2361111	0694444
6	1666667	0142857	3945313	0195313	(2)167411	112675	(2)487076	(3)241127
7	1428571	0357143	3333333	047619	0119048	2625661	0324074	(2)462963
8	125	(2)595238	2890625	(2)78125	(3)372024	0419635	(3)973274	(4)263047
9	1141111	0166667	2554113	021645	(2)324675	1143378	(2)827721	(3)701459
10	1	(2)30303	228906	(2)390625	(3)118371	0204013	(3)296434	(5)50586
11	0909091	(2)909091	2074593	011655	(2)11655	0603794	(2)288138	(3)161875
12	0833333	(2)174825	1897321	(2)223214	(4)468282	0114945	(3)114662	(5)134896
13	0769231	(2)549451	1748252	(2)699301	(3)4995	0358461	(2)121406	(4)485625
14	0714286	(2)10989	1621094	(2)139509	(4)214629	(2)712567	(4)518655	(6)446347
15	0666667	(2)357143	1511312	(2)452489	(3)242405	0230459	(3)583068	(4)174571
16	0625	(3)735294	1415551	(3)93006	(4)109419	(2)4726	(4)262201	(6)172274
17	0588235	(2)245098	1331269	(2)309598	(3)128999	015702	(3)308164	(5)716661
18	0555556	(3)515996	125651	(3)651042	(5)604683	(2)329671	(4)144079	(7)746522
19	0526316	(2)175439	1189739	(2)221141	(4)737137	0111832	(3)175256	(5)325755
20	05	(3)37594	1129735	(3)473485	(5)356004	(2)239172	(5)844838	(7)354081
21	047619	(2)12987	1075515	(2)163452	(4)445778	(2)2824851	(3)10562	(5)160517
22	0454545	(3)282326	1026278	(3)355114	(5)220567	(2)179056	(5)521881	(7)180582
23	0434783	(3)988142	0981366	(2)124224	(4)282326	(2)625908	(4)667207	(6)844567
24	0416667	(3)217391	0940232	(3)273164	(5)142521	(2)137548	(5)336458	(8)977507
25	04	(3)769231	0902415	(3)966184	(4)185805	(2)486235	(4)438236	(6)469203
26	0384615	(3)17094	0867531	(3)214629	(6)953907	(2)107959	(5)224803	(8)556168
27	037037	(3)610501	0835249	(3)766284	(4)12631	(2)385274	(4)297453	(6)272893
28	0357143	(3)136836	0805288	(3)171703	(6)657867	(3)862955	(5)154823	(8)330112
29	0344828	(3)492611	0777407	(3)617971	(5)882815	(2)310473	(4)207641	(6)165056
30	0333333	(3)111235	0751395	(3)139509	(6)465547	(3)700681	(5)10944	(8)203194
31	0322581	(3)403226	072707	(3)505612	(5)632015	(2)25387	(4)148503	(6)103271
32	03125	(4)916422	0704274	(3)11489	(6)33692	(3)576715	(6)79131	(8)129088
33	030303	(3)334225	0682866	(3)418936	(5)462062	(2)210245	(4)10848	(7)665521
34	0294118	(4)763942	0662722	(4)957414	(6)248679	(3)480378	(6)583624	(9)843143
35	0285714	(3)280112	0643735	(3)351	(5)344118	(2)176078	(5)807344	(7)440209
36	0277778	(4)643501	0625806	(4)806244	(6)186775	(3)404376	(6)438065	(9)564371

Таблица 30П. Множители M для четных рядов (нуль целых опущен)

N	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
6	1666667	0571429	3945313	078125	0267857	4507	0779391	0154321
8	125	0238095	2890625	03125	0059524	1678541	0155724	0016835
10	1	0121212	2291233	0156398	0018957	0816053	0047429	0003238
12	0833333	0069930	1897321	0089986	0007493	0459779	0018346	0000863
14	0714286	0043956	1621094	0055804	0003434	0285027	0008298	0000286
16	0625	0029412	1415551	0037202	0001751	0189040	0004195	000011
18	0555556	002064	125651	0026042	0000967	0131869	0002305	0000048
20	05	0015038	1129735	0018939	000057	0095669	0001352	0000023
22	0454545	0011293	1026278	0014197	0000353	0071622	0000835	0000011
24	0416667	0008696	0940232	0010927	0000228	0055019	0000538	0000006

Таблица 31П. Вспомогательные величины для вычисления коэффициентов уравнений парабол способом сумм

N	$\sum x^2$	$\sum x^4$	$\sum x^6$	D_2	D_3
2	0,5	0,125	0,03125	0	0
3	2	2	2	2	0
4	5	10,25	22,8125	16	9
5	10	34	130	70	144
6	17,5	88,375	511,09375	224	1 134
7	28	196	1 588	588	6 048
8	42	388,5	4 187,625	1 344	24 948
9	60	708	9 780	2 772	85 536
10	82,5	1 208,625	20 795,15625	5 280	254 826
11	110	1 958	41 030	9 438	679 536
12	143	3 038,75	76 156,4375	16 016	1 656 369
13	182	4 550	134 342	26 026	3 747 744
14	227,5	6 608,875	226 994,21875	40 768	7 963 956
15	280	9 352	369 640	61 880	16 039 296
16	340	12 937	582 951,25	91 392	30 837 456
17	408	17 544	893 928	131 784	56 930 688
18	484,5	23 377,125	1 337 250,28125	186 048	101 407 788
19	570	30 666	1 956 810	257 754	174 978 144
20	665	39 667,25	2 807 434,0625	351 120	293 452 929
21	770	50 666	3 956 810	471 086	479 700 144
22	885,5	63 977,375	5 487 625,34375	623 392	766 187 730
23	1012	79 948	7 499 932	814 660	1 198 248 480
24	1150	98 957,5	10 113 746,375	1 052 480	1 838 222 100
25	1300	121 420	13 471 900	1 345 500	2 770 653 600

Таблица 32П. Выражение часов, минут и секунд в дуговых мерах

Часы		Минуты			Секунды				
1 ^ч	15°	1 ^м	0°15′	31 ^м	7°45′	1 ^с	0′15″	31 ^с	7′45″
2	30°	2	0 30	32	8 0	2	0 30	32	8 0
3	45	3	0 45	33	8 15	3	0 45	33	8 15
4	60	4	1 10	34	8 30	4	1 0	34	8 30
5	75	5	1 15	35	8 45	5	1 15	35	8 45
6	90	6	1 30	36	9 0	6	1 30	36	9 0
7	105	7	1 45	37	9 15	7	1 45	37	9 15
8	120	8	2 0	38	9 30	8	2 0	38	9 30
9	135	9	2 15	39	9 45	9	2 15	39	9 45
10	150	10	2 30	40	10 0	10	2 30	40	10 0
11	165	11	2 45	41	10 15	11	2 45	41	10 15
12	180	12	3 0	42	10 30	12	3 0	42	10 30
13	195	13	3 15	43	10 45	13	3 15	43	10 45
14	210	14	3 30	44	11 0	14	3 30	44	11 0
15	225	15	3 45	45	11 15	15	3 45	45	11 15
16	240	16	4 0	46	11 30	16	4 0	46	11 30
17	255	17	4 15	47	11 45	17	4 15	47	11 45
18	270	18	4 30	48	12 0	18	4 30	48	12 0
19	285	19	4 45	49	12 15	19	4 45	49	12 15
20	300	20	5 0	50	12 30	20	5 0	50	12 30
21	315	21	5 15	51	12 45	21	5 15	51	12 45
22	330	22	5 30	52	13 0	22	5 30	52	13 0
23	345	23	5 45	53	13 15	23	5 45	53	13 15
24	360	24	6 0	54	13 30	24	6 0	54	13 30
		25	6 15	55	13 45	25	6 15	55	13 45
		26	6 30	56	14 0	26	6 30	56	14 0
		27	6 45	57	14 15	27	6 45	57	14 15
		28	7 0	58	14 30	28	7 0	58	14 30
		29	7 15	59	14 45	29	7 15	59	14 45
		30	7 30	60	15 0	30	7 30	60	15 0

Таблица 33П. Часы, минуты и секунды в долях суток

Часы в долях суток		Минуты в долях суток				Секунды в долях суток	
1	0,041667	1	0,000694	31	0,021528	1	0,000012
2	0,083333	2	0,001389	32	0,022222	2	0,000023
3	0,125000	3	0,002083	33	0,022917	3	0,000035
4	0,166667	4	0,002778	34	0,023611	4	0,000046
5	0,208333	5	0,003472	35	0,024305	5	0,000058
6	0,250000	6	0,004167	36	0,025000	6	0,000069
7	0,291667	7	0,004861	37	0,025694	7	0,000081
8	0,333333	8	0,005556	38	0,026389	8	0,000093
9	0,375000	9	0,006250	39	0,027083	9	0,000104
10	0,416667	10	0,006944	40	0,027778	10	0,000116
11	0,458333	11	0,007639	41	0,028472	11	0,000127
12	0,500000	12	0,008333	42	0,029167	12	0,000139
13	0,541667	13	0,009028	43	0,029861	13	0,000150
14	0,583333	14	0,009722	44	0,030556	14	0,000162
15	0,625000	15	0,010417	45	0,031250	15	0,000174
16	0,666667	16	0,011111	46	0,031944	16	0,000185
17	0,708333	17	0,011805	47	0,032639	17	0,000197
18	0,750000	18	0,012500	48	0,033333	18	0,000208
19	0,791667	19	0,013194	49	0,034028	19	0,000220
20	0,833333	20	0,013889	50	0,034722	20	0,000231
21	0,875000	21	0,014583	51	0,035417	30	0,000347
22	0,916667	22	0,015278	52	0,036111	40	0,000463
23	0,958333	23	0,015972	53	0,036805	50	0,000579
24	1,000000	24	0,016667	54	0,037500	60	0,000694
		25	0,017361	55	0,038194		
		26	0,018055	56	0,038889		
		27	0,018750	57	0,039583		
		28	0,019444	58	0,040278		
		29	0,020139	59	0,040972		
		30	0,020833	60	0,041667		

Таблица 34П. Выражение дуги в часах, минутах и секундах

Градусы в часах, минутах и секундах					Минуты дуги в часах, минутах и секундах					Секунды дуги в часах, минутах и секундах				
1°	0 ^ч	4 ^м	120 ^с	8 ^ч 0 ^м	1'	0 ^м 4 ^с	31'	2 ^м 4 ^с		1"	0 ^с ,07	31"	2 ^с ,07	
2	0 ^ч	8 ^м	130	8 40	2	8	32	8		2	0,13	32	2,13	
3	0	12	140	9 20	3	12	33	12		3	0,20	33	2,20	
4	0	16	150	10 0	4	16	34	16		4	0,27	34	2,27	
5	0	20	160	10 40	5	20	35	20		5	0,33	35	2,33	
6	0	24	170	11 20	6	24	36	24		6	0,40	36	2,40	
7	0	28	180	12 0	7	28	37	28		7	0,47	37	2,47	
8	0	32	190	12 40	8	32	38	32		8	0,53	38	2,53	
9	0	36	200	13 20	9	36	39	36		9	0,60	39	2,60	
10	0	40	210	14 0	10	40	40	40		10	0,67	40	2,67	
11	0	44	220	14 40	11	44	41	44		11	0,73	41	2,73	
12	0	48	230	15 20	12	48	42	48		12	0,80	42	2,80	
13	0	52	240	16 0	13	52	43	52		13	0,87	43	2,87	
14	0	56	250	16 40	14	0 56	44	2 56		14	0,93	44	2,93	
15	1	0	260	17 20	15	1 ^м 0 ^с	45	3 0		15	1,00	45	3,00	
20	1	20	270	18 0	16	4	46	4		16	1,07	46	3,07	
30	2	0	280	18 40	17	8	47	8		17	1,13	47	3,13	
40	2	40	290	19 20	18	12	48	12		18	1,20	48	3,20	
50	3	20	300	20 0	19	16	49	16		19	1,27	49	3,27	
60	4	0	310	20 40	20	20	50	20		20	1,33	50	3,33	
70	4	40	320	21 20	21	24	51	24		21	1,40	51	3,40	
80	5	20	330	22 0	22	28	52	28		22	1,47	52	3,47	
90	6	0	340	22 40	23	32	53	32		23	1,53	53	3,53	
100	6	40	350	23 20	24	36	54	36		24	1,60	54	3,60	
110	7	20	360	24 0	25	40	55	40		25	1,67	55	3,67	
					26	44	56	44		26	1,73	56	3,73	
					27	48	57	48		27	1,80	57	3,80	
					28	52	58	52		28	1,87	58	3,87	
					29	1 ^м 56 ^с	59	3 56		29	1,93	59	3,93	
					30	2 0	60	4 ^м 0 ^с		30	2,00	60	4,00	

ЛИТЕРАТУРА

- Браунли К. А.* Статистическая теория и методология в науке и технике. М.: Наука, 1977. 407 с.
- Бреев К. А.* Применение негативного биномиального распределения для изучения популяционной экологии паразитов. Л.: Наука, 1972. 73 с.
- Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе. М.: Статистика, 1974. 279 с.
- Вентцель Е. С.* Теория вероятностей. М.: Наука, 1969. 576 с.
- Воробьев Н. Н.* Числа Фибоначчи. М.: Наука, 1969. 110 с.
- Доерфель К.* Статистика в аналитической химии. М.: Мир, 1969. 247 с.
- Воронцов-Вельяминов Б. А.* Сборник задач и практических упражнений по агрономии. 7-е изд. М.: Наука, 1977. 262 с.
- Дружинин Н. К.* Математико-статистические методы анализа экспериментальных данных в товароведении. М.: Гос. изд-во торговой литературы, 1959. 164 с.
- Зайцев Г. Н.* Методика биометрических расчетов. М.: Наука, 1973. 256 с.
- Зайцев Г. Н.* Фенология травянистых многолетников. М.: Наука, 1979. 150 с.
- Зайцев Г. Н.* Фенология древесных растений. М.: Наука, 1981. 120 с.
- Кендалл М., Стьюарт А.* Теория распределений. М.: Наука, 1966. 587 с.
- Комаров В. Л.* Монография рода *Сагана*.— Тр. Имп. СПб. ботан. сада. СПб. 1908, т. 29.
- Ляхтин Л. К.* Кривые распределения и построение для них интерполяционных формул по способу Пирсона и Брунса. М.: ГИЗ, 1922.
- Леонтьев Н. Л.* Техника статистических вычислений. М.: Гослесбумиздат, 1966. 250 с.
- Липкин М. И.* Кривые распределения в экономических исследованиях. М.: Статистика, 1972. 144 с.
- Митропольский А. К.* Техника статистических вычислений. М.: Наука, 1971. 576 с.
- Пановский Г. А., Брайер Г. В.* Статистические методы в метеорологии. Л.: Гидрометеиздат, 1967. 242 с.
- Плохинский Н. А.* Биометрия. Новосибирск. Изд-во АН СССР, 1961. 364 с.
- Плохинский Н. А.* Математические методы в биологии. М.: Изд-во МГУ, 1978. 265 с.
- Поморский Ю. Л.* Вариационная статистика. Л., 1930, ч. II. 463 с.
- Пустыльник Е. И.* Статистические методы анализа и обработки наблюдений. М.: Наука, 1968. 288 с.
- Рокицкий П. Ф.* Биологическая статистика. Минск: Высшая школа, 1967. 327 с.
- Романовский В. И.* Применения математической статистики в опытном деле. М.: Гостехиздат, 1947. 247 с.
- Селиванов Н. А.* Математические методы в собирании и исследовании доказательств. М.: Юридическая литература, 1974. 120 с.
- Семендяев К. А.* Эмпирические формулы. М.; Л.: Гос. техн.-теор. изд-во, 1933. 88 с.
- Сепетлиев Д.* Статистические методы в научных медицинских исследованиях. М.: Медицина, 1968. 419 с.
- Смирнов Е. С.* Таксономический анализ. М.: Изд-во МГУ, 1969. 187 с.
- Смирнов Н. В., Дунин-Барковский И. В.* Курс теории вероятностей и математической статистики для технических приложений. Изд. 3. М.: Наука, 1969. 511 с.

- Снедекор Д. У.* Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. М.: Сельхозиздат, 1961.
- Терентьев П. В., Ростова Н. С.* Практикум по биометрии. Л., Изд-во ЛГУ, 1977. 152 с.
- Урбах В. Ю.* Математическая статистика для биологов и медиков. М.: Изд-во АН СССР, 1963. 323 с.
- Хан Г., Шапиро С.* Статистические модели в инженерных задачах. М.: Мир, 1969. 395 с.
- Хастингс Н., Пикок Дж.* Справочник по статистическим распределениям. М.: Статистика, 1980. 96 с.
- Хотимский В. И.* Выравнивание статистических рядов по методу наименьших квадратов (способ Чебышева). М.: Госстатиздат, 1959. 88 с.
- Четвериков Н. С.* Статистические исследования (теория и практика). М.: Наука, 1975. 388 с.
- Шторм Р.* Теория вероятностей. Математическая статистика. Статистический контроль качества. М.: Мир, 1970. 368 с.
- Юл Д. Э., Кендалл М. Д.* Теория статистики. М.: Госстатиздат, 1960. 779 с.
- Ястремский Б. С.* Избранные труды. М.: Статистика, 1964. 391 с.
- Martin R.* Lehrbuch der Anthropologie in systematischer Darstellung mit besonderer Berücksichtigung der anthropologischen Methoden. 2. Aufl. Jena: Verl. Fischer, 1928. 1. Bd.
- Weber E.* Grundriss der biologischen Statistik. 6. Aufl. Jena: Verl. Fischer, 1967. 674 S.

Таблицы

- Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. М.: ВЦ АН СССР, 1968. 474 с. (Библ. мат. табл. Вып. 46).
- Келли Т. Л.* Статистические таблицы. М.: ВЦ АН СССР, 1966. 194 с. (Библ. мат. табл. вып. 35).
- Митропольский А. К.* Статистическое исчисление. Л.: Изд. ВЗЛТИ, 1952—1954, вып. 1—4. 166 с.; 191 с.; 200 с.; 200 с.
- Митропольский А. К.* Краткие математические таблицы. М.: Наука, 1968. 95 с.
- О'Рурк.* Таблицы умножения. М., Госстатиздат, 1962. 336 с.
- Сегал Б. И., Семендяев К. А.* Пятизначные математические таблицы. 3-е изд. М.: Гос. изд. физ.-мат. лит., 1962. 464 с.
- Таблицы Барлоу/Под ред. и доп. проф. Л. С. Хренова. М.: Мир, 1965. 326 с.
- Янко Я.* Математико-статистические таблицы. М.: Госстатиздат, 1961. 243 с.
- Pearson K.* Tables for statisticians and biometricians. 3rd ed. L., 1931, Pt. II.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
Глава 1. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ВАРИАЦИОННОГО РЯДА	5
§ 1.01. Составление вариационных рядов	5
§ 1.02. Оценка сильно отклоняющихся вариантов	10
§ 1.03. Средняя арифметическая невзвешенного ряда. Уровни значимости	11
§ 1.04. Средняя арифметическая взвешенного ряда	14
§ 1.05. Общая средняя арифметическая по частным средним	17
§ 1.06. Средняя гармоническая	17
§ 1.07. Средняя квадратическая и средняя кубическая	18
§ 1.08. Средняя геометрическая	21
§ 1.09. Среднее квадратическое отклонение невзвешенного ряда	22
§ 1.10. Среднее квадратическое отклонение взвешенного ряда	26
§ 1.11. Приближенное вычисление среднего квадратического отклонения по способу случайных серий, размаху варьирования и коэффициенту вариации	33
§ 1.12. Объединение выборок по их дисперсиям	34
§ 1.13. Коэффициент вариации	35
§ 1.14. Показатель точности опыта	37
§ 1.15. Медиана	38
§ 1.16. Вычисление 11 параметров невзвешенного ряда	39
§ 1.17. Мода	41
§ 1.18. Условные, центральные и факториальные моменты распределения	42
§ 1.19. Асимметрия	55
§ 1.20. Эксцесс	57
§ 1.21. Энтропия	59
§ 1.22. Вычисление параметров взвешенного ряда	61
§ 1.23. Доверительные интервалы статистических параметров	65
§ 1.24. Определение репрезентативности и методы отбора вариантов_выборок	67
§ 1.25. Нормальное распределение	72
§ 1.26. Альтернативное распределение	79
§ 1.27. Биномиальное распределение	82
§ 1.28. Гипергеометрическое распределение	88
§ 1.29. Распределение Пуассона	92
§ 1.30. Негативное биномиальное распределение	96
§ 1.31. Мультиномиальное распределение	104
§ 1.32. Распределение Максвелла	105
§ 1.33. Распределение Рэля	110
§ 1.34. Показательное распределение	112
§ 1.35. Распределение Парето	118
§ 1.36. Равномерное распределение	121
§ 1.37. Логнормальное распределение	124

Глава 2. КОРРЕЛЯЦИЯ	130
§ 2.01. Корреляционная решетка и эмпирическая линия регрессии	132
§ 2.02. Показатель корреляции рангов по Спирмэну	135
§ 2.03. Коэффициент корреляции рангов по Кендэлу	138
§ 2.04. Коэффициент корреляции для малых выборок	142
§ 2.05. Коэффициент корреляции для больших выборок	145
§ 2.06. Совместное вычисление коэффициента корреляции и прямого корреляционного отношения между взвешенными рядами	149
§ 2.07. Критерии криволинейности.	152
§ 2.08. Прямое корреляционное отношение для невзвешенных рядов	154
§ 2.09. Обратное корреляционное отношение для невзвешенных рядов	158
§ 2.10. Прямое корреляционное отношение для взвешенных рядов	159
§ 2.11. Обратное корреляционное отношение для взвешенных рядов	161
§ 2.12. Полихорический коэффициент сопряженности Чупрова	162
§ 2.13. Тетрахорический коэффициент сопряженности	164
§ 2.14. Бисериальный коэффициент корреляции	166
§ 2.15. Бисериальный коэффициент корреляции по таблице Келли—Вуда	168
§ 2.16. Частные и множественные коэффициенты корреляции	170
Глава 3. РЕГРЕССИЯ	175
§ 3.01. Уравнение прямой линии	175
§ 3.02. Уравнение множественной регрессии	186
§ 3.03. Парабола второго порядка	193
§ 3.04. Парабола третьего порядка	199
§ 3.05. Обращенная парабола	203
§ 3.06. Вычисление коэффициентов уравнений парабол способом сумм	209
§ 3.07. Степенная функция	218
§ 3.08. Показательно-степенная функция	221
§ 3.09. Показательная функция	223
§ 3.10. Показательное уравнение второго порядка	225
§ 3.11. Логистическая функция	227
§ 3.12. Кривая Гаусса	237
§ 3.13. Логарифмическая функция	240
§ 3.14. Метод избранных точек	247
Глава 4. СРАВНЕНИЕ ВЫБОРОК И СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ	251
§ 4.01. Сравнение средних арифметических и дисперсий	251
§ 4.02. Парный критерий Стьюдента	253
§ 4.03. Сравнение коэффициентов вариации и показателей точности опыта	255
§ 4.04. Сравнение долей из больших выборок	257
§ 4.05. Сравнение долей по методу Фишера	258
§ 4.06. Сравнение коэффициентов корреляции	259
§ 4.07. Критерий лямбда	260
§ 4.08. Критерий медианы	262
§ 4.09. Критерий Вилкоксона	264
§ 4.10. Критерий ван дер Вардена	266
§ 4.11. Серийный критерий	267

§ 4.12.	Критерий знаков	268
§ 4.13.	Парный критерий Вилкоксона	269
§ 4.14.	Критерий хи-квадрат	270
§ 4.15.	Сравнение частот взвешенных рядов по критерию Колмогорова	283
§ 4.16.	Двухфакторный дисперсионный анализ с повторностями	286
§ 4.17.	Однофакторный дисперсионный анализ небольшой группы данных	294
§ 4.18.	Двухфакторный дисперсионный анализ небольшой группы данных без повторностей	299
§ 4.19.	Двухфакторный дисперсионный анализ большой группы данных	302
§ 4.20.	Критерий Бартлетта	308
§ 4.21.	Ранговый критерий Фридмана	312
§ 4.22.	Ранговый критерий Крускала—Уоллиса	315
Глава 5.	СПЕЦИАЛЬНЫЕ БИОЛОГО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ	317
§ 5.01.	Построение шкал балльной оценки	317
§ 5.02.	Кодирование признаков жизненных форм высших растений	331
§ 5.03.	Таксономический анализ Смирнова	333
§ 5.04.	Сравнение флор по видовому составу	339
§ 5.05.	Логистический анализ	343
§ 5.06.	Пробит-анализ	352
§ 5.07.	Применение показателей атипичности и аномальности для идентификации особи по комплексу признаков	356
§ 5.08.	Критерий пропорциональности в ландшафтной архитектуре	358
§ 5.09.	Усреднение и нормирование взвешенных вариационных рядов	369
§ 5.10.	Задачи на составление смесей	372
§ 5.11.	Определение непрístupных расстояний	375
§ 5.12.	Формулы перевода результатов измерений температуры и других градусных мер	380
§ 5.13.	Определение площади сосудов на срезе стебля	381
§ 5.14.	Комплексная оценка надежности результатов массовых фенологических наблюдений	382
ПРИЛОЖЕНИЯ		388
Литература		420

Геннадий Николаевич
ЗАЙЦЕВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА в экспериментальной ботанике

Утверждено к печати Главным ботаническим садом Академии наук СССР.

Редактор издательства Л. К. Соколова. Художник А. С. Александров.

Художественный редактор М. В. Версоцкая.

Технический редактор О. М. Гуськова. Корректор А. Б. Васильев.

ИБ № 28031

Сдано в набор 16.03.84. Подписано к печати 14.11.84. Т-19877. Формат 60×90¹/₁₆. Бумага книжно-журнальная. Гарнитура обыкновенная. Печать высокая. Усл. печ. л. 26,5. Уч.-изд. л. 29,7. Усл. кр. отт. 26,5. Тираж 2800 экз. Тип. зак. 300. Цена 3 р. 30 к.

Издательство «Наука» 117864 ГСП-7, Москва В-485 Профсоюзная ул., 90]

2-я типография издательства «Наука» 121099, Москва, Г-99, Шубинский пер., 10